

Saidha, Tekchand

From: Saidha, Tekchand
Sent: Monday, June 23, 2003 2:47 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request - 09/830751

09/830751

Please search the data base and interference files for :

SEQ ID Nos. 2, 4, 6 and 8

Thank you !

Tekchand Saidha

Primary Examiner

Art Unit 1652, CM1, Room No. 10D05

Mail Box 10D01

(703) 305-6595





STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number

TO: Tekchand Saidha
Location: CM1/10D05/10D01
Art Unit: 1652
Tuesday, June 24, 2003

Case Serial Number: 830751

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Saidha,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart



97247

STIC-Biotech/ChemLib

From: Saidha, Tekchand
Sent: Monday, June 23, 2003 2:47 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request - 09/830751

09/830751

Please search the data base and interference files for :

SEQ ID Nos. 2, 4, 6 and 8

Thank you !

Jekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595

CRIFE

6/24/03
ABSSQ2
4-ATT





STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1—Circ Desk



18 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
19 fermenting recombinant microorganisms expressing genes for suitable
20 enzymes in the presence of glycerol or glucose -
21 claim 5, Page 26-28, 64pp, English.

22 The present invention describes a method for the production of
23 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
24 recombinant microorganism in the presence of a source of glycerol (1)
25 or glucose, where the microorganism: (i) expresses genes for non-native
26 enzymes which catalyse the production of (3-HP) from (1); (ii) carries
27 genetic constructions for the expression of a glycerol dehydratase
28 (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
29 production of (3-HP) from (1); or (iii) carries a genetic construct
30 which expresses the *dhap* gene from *Klebsiella pneumoniae* and a gene for
31 an ADH capable of catalysing the production of (3-HP) from (1). 3-HP is
32 a monomer, and is useful e.g. in the production of absorbable prosthetic
33 devices and surgical sutures or for incorporation into beta-lactams,
34 production of acyclic acid or formation of trifluoromethylated alcohols
35 or diols, polyhydroxyalkonates and copolymers with lactic acid.
36 Incorporation of genes encoding two enzymes makes the host organisms
37 able to produce (3-HP) from (1). The biotechnological method of
38 preparing (3-HP) is potentially cheaper than chemical synthesis. The
39 present sequence represents the yeast aldehyde dehydrogenase Ald4, which
40 is used in the exemplification of the present invention.

41 Sequence: 495 AA;

42 Query Match: 100.0%; Score 2550; DB 22; Length 495;
43 Best Local Similarity: 100.0%; Pred. No. 1.7e-207;

44 Matches: 495; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

45 1 MSHLPMTVPKIKLNGLEFEGTLEINNKTPSKUNKTFEVLNPSLEELCHLYEGRED 60
46 1 MSHLPMTVPKIKLNGLEFEGTLEINNKTPSKUNKTFEVLNPSLEELCHLYEGRED 60
47 61 VEFVQADAPAFNSGNNIPIDRCKALYRIAEHLIYKQKDVIASTETLDNKAISSSG 120
48 61 VEFVQADAPAFNSGNNIPIDRCKALYRIAEHLIYKQKDVIASTETLDNKAISSSG 120
49 121 IVDVIVNLEKSSAGFAIKIKKRMIDGRTHSYTKKQPLGVCGLIPMNPFLIMWAKTA 180
50 121 IVDVIVNLEKSSAGFAIKIKKRMIDGRTHSYTKKQPLGVCGLIPMNPFLIMWAKTA 180
51 181 PALVTGNTVTKTAESETHSALYVSKYTPQAGIPPCVINIVSGEKKIVEAATINPKIK 240
52 181 PALVTGNTVTKTAESETHSALYVSKYTPQAGIPPCVINIVSGEKKIVEAATINPKIK 240
53 241 VAFGSGTATGRIHLYSAAAGIKKVTLELGGKSSPNIVFADAEIKKAVQNTILGTYNSGEV 300
54 241 VAFGSGTATGRIHLYSAAAGIKKVTLELGGKSSPNIVFADAEIKKAVQNTILGTYNSGEV 300
55 401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
56 401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
57 401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
58 401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
59 461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
60 461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
61 461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
62 461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
63 421 DSEFGTAAAGIHTSNTINIALKVAQVAVANGITWINTYNIPIHIAVPGVGFNSAGLREMSVIA 480
64 421 DSEFGTAAAGIHTSNTINIALKVAQVAVANGITWINTYNIPIHIAVPGVGFNSAGLREMSVIA 480
65 481 LGNTLVKAVAKID 495
66 481 LGNTLVKAVAKID 495

67 RESULT 2
68 AAF60456 standard: Proteins: 497 AA.
69 XX
70 XX

71 AAF60456.
72 25-JUN-1991 (first entry)
73 XX
74 XX
75 XX
76 XX
77 XX
78 XX
79 XX
80 XX
81 XX
82 XX
83 XX
84 XX
85 XX
86 XX
87 XX
88 XX
89 XX
90 XX
91 XX
92 XX
93 XX
94 XX
95 XX
96 XX
97 XX
98 XX
99 XX
100 XX

Sequence of aldehyde dehydrogenase (ald4) of *Aspergillus nidulans*.

Filamentous fungi promoter.

Aspergillus nidulans.

W08506097-A.

23-OCT-1986.

14-APR-1986; 86W0-GB00209.

20-DEC-1985; 85US-0811404.

15-APR-1985; 85CA-0479135.

(ALF-) ALLELIX INC.

Gwynne D.I., Buxton F., Pickett M., Davies R., Sczarcchio C;

WPI: 1986-291664/44.

N-PSDB: AAN60401.

DNA construct for use in filamentous fungi - comprising promoter

operative in filamentous fungi to promote transcription of coding

region

disclosure; Fig 1A; 75pp; English.

In the constructs of the invention, the promoter region naturally
associated with the alcohol dehydrogenase I (*ald4*) gene and the
aldehyde dehydrogenase (*ald4*) gene of *A. nidulans* or naturally
associated with the glucanase gene in *Aspergillus niger* may be
used. The DNA construct may contain a promoter region in operative
association with a signal peptide coding region. The promoter/signal
construct is suitably provided with a flanking restriction site to
allow precise coupling of the protein coding region to the signal
peptide coding region.

Sequence: 497 AA;

Query Match: 55.1%; Score 1405.5; DB 7; Length 497;

Best Local Similarity: 54.3%; Pred. No. 1.9e-110;

Matches: 269; Conservative: 88; Mismatches: 133; Indels: 5; Gaps: 3;

1 MSHLPMTVPKIKLNGLEFEGTLEINNKTPSKUNKTFEVLNPSLEELCHLYEGRED 60
1 MSHLPMTVPKIKLNGLEFEGTLEINNKTPSKUNKTFEVLNPSLEELCHLYEGRED 60
61 VEFVQADAPAFNSGNNIPIDRCKALYRIAEHLIYKQKDVIASTETLDNKAISSSG 120
61 VEFVQADAPAFNSGNNIPIDRCKALYRIAEHLIYKQKDVIASTETLDNKAISSSG 120
121 IVDVIVNLEKSSAGFAIKIKKRMIDGRTHSYTKKQPLGVCGLIPMNPFLIMWAKTA 180
121 IVDVIVNLEKSSAGFAIKIKKRMIDGRTHSYTKKQPLGVCGLIPMNPFLIMWAKTA 180
181 PALVTGNTVTKTAESETHSALYVSKYTPQAGIPPCVINIVSGEKKIVEAATINPKIK 240
181 PALVTGNTVTKTAESETHSALYVSKYTPQAGIPPCVINIVSGEKKIVEAATINPKIK 240
241 VAFGSGTATGRIHLYSAAAGIKKVTLELGGKSSPNIVFADAEIKKAVQNTILGTYNSGEV 300
241 VAFGSGTATGRIHLYSAAAGIKKVTLELGGKSSPNIVFADAEIKKAVQNTILGTYNSGEV 300
401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
401 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 360
461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
461 NEBNTLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 420
421 DSEFGTAAAGIHTSNTINIALKVAQVAVANGITWINTYNIPIHIAVPGVGFNSAGLREMSVIA 480
421 DSEFGTAAAGIHTSNTINIALKVAQVAVANGITWINTYNIPIHIAVPGVGFNSAGLREMSVIA 480
481 LGNTLVKAVAKID 495
481 LGNTLVKAVAKID 495
300 WCCAGSRVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 359
300 WCCAGSRVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 359
297 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 356
297 CTAGSPVYVEESTYDKPIPEPKAASPIKVGDPFDESTFGAGTSSQMLNKLTKYVDIK 356
360 KNPQATLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 419
360 KNPQATLITGGERLSKSYFLEKTVGIVKEDMEYVEFTPGPVVYVTKRKSANFVIMAN 419

DB 357 KAGAGTAVAGDGRHENEVPIQPPVPTIVTISDMKTAQFIPGPVVIQKREDAVAFAKIG 416
 420 NSESEGLAAGHTSINIALKADRVNAGTWINIYMDPHHAAVPGSEFNASGLSKREMSVD 479
 417 XSTQGLAAAVITKRNVAIAIKVSNALKAGIYWINNNYMSVQAPFGPKSGSLCPREIGSY 476
 QY 480 ALONTLOKAVPAKL 494
 477 ALENYTOIKIVHYRL 491

RESULT 3

AAK71891
 ID AAK71891 standard. Protein: 496 AA

AC AAK71891:
 DT 25-OCT-1995 (first entry)

DE Cladosporium herbarum allergen ClaH53.

KW Fungal spore, allergen, ClaH53, allergy: aldehyde dehydrogenase.

OS Cladosporium herbarum

PN W09506121-A.

PD 02-MAR-1995.

PF 24-AUG-1994: 94W0-AT90120.

PR 27-AUG-1993: 93AT-0001725.

PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.

PI Achatz G., Breitenbach M., Ebner C., Hirschwehr R.

PI Kraft D., Lechenauer E., Oberkofler H., Prillinger H.

PI Simon B., Unger A.

DR WPI: 1995-106850/14.

DR N-PSDB: AA086278.

PT Allergens derived from Cladosporium herbarum spores - also

PI recombinant DNA for expressing the allergens, useful for in vitro

PI allergy detection

PS Claim 1, Page 8 3, 35pp; German.

CC Spores of Cladosporium herbarum are the most common fungal spores

CC found in the air; they can cause allergic reactions. Various ClaH

CC allergens and sequences encoding them have now been isolated. The

CC mature ClaH53 allergen has mol. wt. 53 kD and is encoded by cDNA

CC sequence AA086278. The allergen has homology to aldehyde

CC dehydrogenases. Potential epitopic subfragments were identified by

CC computer analysis of the amino acid sequence. See AAK71892-471906 for

CC potential B-cell epitopes and AAK72615-472627 for potential T-cell

CC epitopes.

XX Sequence 496 AA;

Query Match 54 9%; Score 1399, DR 16, length 496;

Best Local Similarity 53.8%; Pred. No. 6, 7e-110;

Matches 263; Conservative 89; Mismatches 155; Indels 2; Gaps 2.

DB 7 TTPPIKIPNIGYPPQPTGIFITNNKVPVSKONTPEVINPDSITFEITGHIVYERENDVPERAVD 66

DB 3 SVQTEIPHSKYKPEFTQIFINNEFYKGEKTFEIVINSDESVILQVHEAMIEKLVDAVA 62

QY 67 AADRAFSNSWNGITPTPEKATVYAEIIFGKIVIASITLNNKKAISSSKQVNLVI 126

DB 63 AAKQAF-ETSMPIETFPNKKLNNIANNFKNTIDLLAAVSHIDNNKKAISMAVTSN'AS 121

QY 127 NTKSSAPDARKITGGMTITGTHSTYKRGPLGAGVQDIIIPWNPILIMAMKITAALVTS 186
 DB 122 GLLPYVQSMALKITISKVITDITPTDFENVKKEPLSWPSEDSLETPLIMAMKIGPAIAG 181
 QY 187 NTVEKIKESPLSALVYSKFIPLQAGTPEVINIVSGEEKIVEEATINHKIKVAFETS 246
 DB 182 NTIVLKTAMQPTPLGLVAASLVKAGAPPGVINYISGFGVACAAALSSHDVQKVAFTGS 241
 QY 247 TATPEHIVQSA-AALKKVTLFEGSEKSPNIVFAVDELKRAVNNITGIYNSSTEVVAGS 305
 DB 242 TVVQPTILKAAASSNLKAVTLELQSKSPNIVFAVDELKRAVNNITGIYNSSTEVVAGS 301
 QY 306 PVYVPESTYKPEKPEKASNSIKVQPPRPSSTFYVAGTQMDI NKIKVQDQKPPAT 365
 DB 302 PVYVPESTYKPEKPEKASNSIKVQPPRPSSTFYVAGTQMDI NKIKVQDQKPPAT 361
 QY 366 PVYVPESTYKPEKPEKASNSIKVQPPRPSSTFYVAGTQMDI NKIKVQDQKPPAT 425
 DB 362 VETGSEKQDKQYLETETPEVNEEDMKIVKELTFGVGSLATFTYTAIPIYQASVTS 421
 QY 425 LAAGHTSINIALKADRVNAGTWINIYMDPHHAAVPGSEFNASGLSKREMSVDALONT 485
 DB 422 LAAAVHTKNNLIALVSNALKAGIYWINNNYMSVQAPFGPKSGSLCPREIGDALANT 481
 QY 485 QYKAVPAKL 494
 DB 482 QTKIVSIRL 490

RESULT 4

AAK71803
 ID AAK71803 standard. Protein: 495 AA.

AC AAK71803:

DT 02-NOV-1995 (first entry)

DE Alternaria alternata allergen Alta53.

KW Fungal spore, allergen, Alta53, allergy: aldehyde dehydrogenase.

OS Alternaria alternata.

PN W09506122-A.

PD 02-MAR-1995.

PF 24-AUG-1994: 94W0-AT90121.

PR 27-AUG-1993: 93AT-0001725.

PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.

PI Achatz G., Breitenbach M., Ebner C., Hirschwehr R.

PI Kraft D., Lechenauer E., Oberkofler H., Prillinger H.

PI Simon B., Unger A.

DR WPI: 1995-106851/14.

DR N-PSDB: AA086275.

PT Allergens derived from Alternaria alternata - their isolation by

PI reacting sera from patients with fungal extracts; useful for

PI in-vitro allergy detection.

PS Claim 7, Page 8; 26pp; German.

CC Species of Alternaria alternata can cause allergic reactions. Various

CC AltA allergens and sequences encoding them have now been isolated. The

CC mature AltA53 allergen has mol. wt. 53 kD and is encoded by cDNA

CC sequence AA086275. The allergen has homology to aldehyde

CC dehydrogenases. Potential epitopic subfragments were identified by

CC computer analysis of the amino acid sequence. See AAK71804-471816 for

CC potential B-cell epitopes and AAK71817-471832 for potential T-cell

PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/74
 DR N-PSDB; AAS70764.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 36936; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG40377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note. The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 529 AA;

Query Match 47.8%; Score 1218.5; DB 22; Length 529;
 Best Local Similarity 50.4%; Prod. No. 150-94;
 Matches 240; Conservative 84; Mismatches 149; Indels 3; Gaps 3;
 QY 22 TGLFINNKVPKSKNTFEVINSTEETEECHIVYEGPVEFAVGAADPAFNSGSG-WNGI 80
 DB 49 TKIPINNEWHESKSGKFKATCNSTREQICFVFEQKPKPVDKAVEAQAQVAFQFGSPWPL 108
 QY 81 DPTDRCALYPLAETFEFQFVWIASFETENGFATSSSPG IVDLVINLKSSAFADKI 139
 DB 109 DALSRGRILHQADLVERDQATLALETMDTKPFLHAFIDLEGICRTLRVFAQWADKI 168
 QY 140 DGRMDTGRTHFSYTKRQPLGVGCGIIPWNPFLMMWAKIAPALVTGNTVILKTAESTPL 199
 DB 169 QCKTIPDDNVVFTFHPPIGVGSAITPWNFPLMLVWKLAFALCGNTMWLKPATQPL 228
 QY 200 SALYVSKYIQAGIIPRWIVINSGRGIIVVEALTNHPKIKKVAFGTSTATGHHYQSAA- 258
 DB 229 TALYVLSLKEAGFPCCVNVIVPGPPTVGAALISSHPQINKIAFTGCTFVGKLVKEAASR 288
 QY 259 AGLKKVTLLEGGKSPNIVDAELKKAVONIILGIYNSGEVCCAGSRVYFESTYDKFI 318
 DB 289 SNLKRVTLLEGGKSPNIVDAELKKAVONIILGIYNSGEVCCAGSRVYFESTYDKFI 348
 QY 319 HEFKAASES IKGVDPEDESPFGAQTSMQNLKILKYVDIGKNEGATILITGGERIGSKGY 378
 DB 349 RRSVEYAKRPVGDPEVDKTEQGPQIDQKQPKILELTFSGKKEGAKLECGSGAMEDKGL 408
 QY 379 FIKPTVGVKEDMRIVKEIFGPPVVTYTKFSADEVINMANSDEYGLAAGIHTSNINTA 438
 DB 409 FIKPTVSEVIDNMRITAKEEIFGPPVVTYTKFSADEVINMANSDEYGLAAGIHTSNINTA 468
 QY 439 LKLVADRVNACTVWINTYNDHFHVPFGFNASCLGRKMSVIALQNTLOVKAVRAKLI 494
 DB 469 LKLSALESCTVWINCYNALYQAQAPGGGPKMSNGREIGEYALAEYTEYKVTIKL 524

RESULT 11
 ABB60140
 ID ABB60140 standard; Protein: 520 AA.
 XX ABB60140;
 AC ABB60140;
 XX 25-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID No 7212.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS W200171042-A2.
 IN 27-SEP-2001.
 XX 23-MAR-2001; 2001WG-0809241.
 XX 23-MAR-2000; 20000S-191647P.
 PR 11-JUL-2000; 20000S-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PW, Myers RW;
 PI WPI: 2001-656860/75.
 DR N-PSDB; ABL0424.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID No 7212; 21pp + Sequence List; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL0176 ABL0611), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABE7737-ABE72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 520 AA;

Query Match 47.2%; Score 1204.5; DB 22; Length 520;
 Best Local Similarity 50.4%; Prod. No. 270-94;
 Matches 240; Conservative 82; Mismatches 150; Indels 3; Gaps 3;
 QY 22 TGLFINNKVPKSKNTFEVINSTEETEECHIVYEGPVEFAVGAADPAFNSGSG-WNGI 80
 DB 40 TGVFINNEWHESKSGKFIETINFTAEVIAETQALKEIDIDIVAAAKNAKIKLASWGRM 99
 QY 81 DPTDRCALYPLAETFEFQFVWIASFETENGFATSSSPG IVDLVINLKSSAFADKI 139
 DB 100 BASPFGALYPLAETFEFQFVWIASFETENGFATSSSPG IVDLVINLKSSAFADKI 159
 QY 140 DGRMDTGRTHFSYTKRQPLGVGCGIIPWNPFLMMWAKIAPALVTGNTVILKTAESTPL 199
 DB 160 HCKTIPDDNVVFTFHPPIGVGSAITPWNFPLMLVWKLAFALCGNTMWLKPATQPL 219
 QY 200 SALYVSKYIQAGIIPRWIVINSGRGIIVVEALTNHPKIKKVAFGTSTATGHHYQSAA- 258
 DB 229 TALYVLSLKEAGFPCCVNVIVPGPPTVGAALISSHPQINKIAFTGCTFVGKLVKEAASR 278


```

Lb 398 ATLQAGSDKLSKGYIQTPTFSDVKIDMLIATDFIFGPVOTILKFKDLDEVIARANSR 457
QY 424 YCLAAGIHTSINTALKAVDRVAGTWTINTYNDHFHVAHPFGFNRASGLGRMSVDALON 483
Db 458 YGLAAGVFTONLDTAHLDMRALRVGTWVINCDFVLDASIPFGGKMGSGIGREKGYISLNN 517
QY 484 YIQVKAVRAKL 494
Db 518 YIQVKAVVTSL 528

RESULT 15
AAR63672
ID AAR63672 standard; Protein: 521 AA.
XX
AC AAR63672;
XX
DT 08-MAY-1995 (first entry)
XX
DE Aldehyde dehydrogenase Aldh-1.
XX
KW Retro virus; vector, aldehyde-dehydrogenase;
KW glutamylcysteine synthetase; hematopoietic cell, cyclophosphamide;
KW chemotherapy, transgenic animal, gene therapy, cancer therapy;
KW selectable marker.
XX
OS Homo sapiens.
XX
PN W09423015-A.
XX
PD 13-OCT-1994.
XX
PF 01-APR-1994; 94WO-US03624.
XX
PR 01-APR-1993; 93US-0041722.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Dalla-Favera R, Gianni AM;
XX
DR WPI: 1994-333177/41.
DR N-PSDB; AAQ72450.
XX
PT Retroviral vectors encoding human cytosolic aldehyde
PT dehydrogenase or glutamyl cysteine synthetase - used to transform
PT a subject's hematopoietic cells to reduce the toxic effects of
PT cyclo phosphamide chemotherapy
XX
PS Disclosure; Fig.4: 92pp; English.
XX
CC A novel retro virus vector encodes human cytosolic aldehyde-
CC dehydrogenase and/or human glutamylcysteine-synthetase (AAR63673).
CC Hematopoietic cells transfected by the vector are resistant to
CC cyclophosphamide, providing a means of gene therapy that allows
CC higher doses of toxic drugs to be used in cancer chemotherapy.
CC The human genes may also be used as selectable markers for
CC mammalian cell transfection and for transgenic animal breeding.
XX
SQ Sequence 521 AA;

Query Match 45 28; Score 1153 5; DB 15; Length 521;
Best Local Similarity 48.5%; Pred. No. 4.7e-89;
Matches 234; Conservative 88; Mismatches 153; Indels 7; Gaps 5;

QY 20 QPTGLFINNKFPVSKQNKTEVINPSTEEIEICHIEGKELDVEAVQAADRAFSNGS-WN 78
Db 35 QYTKIFINNEHDSVSGKSPFVNPATEELCOVEGSKEDVDKAVKARQAFQIGSPWR 94
QY 79 GIDPIDRKALYRIAELEQTK--DVITASTETLDNGKATSSS--PGVDVLVINYLKSSAGF 135
Db 95 TMDASEPGGLLYKLADLIERDLIATMESMGKLYSNAYLNDLACGCTKTLRYCAGW 154
QY 136 AUKL--DGRMLIDIGKTHESYLKRPGLGVGQQIIPWNPFLLMWAKIAPALVTGNIVLKT 193

```

```

Db 155 ADKLQGGSPITIPIDGNEFFIYTHKPKGVGCGQLLPWNEFLVMLIKIGPAISGNIWVKP 214
QY 194 AESTPI:SAIVVSKYIPGAGTPPQVNTIVSGRPIVVVEATTNHPKIFKVAFTGSTATGRHI 253
Db 215 EEQTPITLHVASLILKEAGFPFGVNVILPGYGPPTAGAAISSHMDIDKVAFTGSTFVWKL 274
QY 254 YQSAA-AGLKKVILELGGKSNIVFAUAEILKKAQVNTILGIYVNSGEVCCAGSRVIVERS 312
Db 275 KEAAGKSNLRVITLLEUGGKSPCIVLADLDNNAVEFAHGVFYHQGCCIAASRIFVERS 334
QY 313 IYDKFIEEPKAASESIKVGDPDFDESTFOGAQTSOMQLNKILKYVDIGKNEGATLITGSHR 372
Db 335 IYDEFVPPSVERAKKYTLGNPLTPGVTOGQPIDKEQYDKILDLESKKKGALECGGGP 394
QY 373 LGSKGYEITPTVFCQVKEQMPVIVKEEIEGPPVTVTVYFKSADENVINMANDEYGLAACIHT 432
Db 395 WGNKGYEQPTVFSNVTDENRIAKBEIEFGVQOIMKFKSLDDVVIKKNNTFYGLSAGVFT 454
QY 433 SNINTALKVALHPVNAQIVWINTYNGPHHVPFGEFNASGEGFPMSVIATQNYLQVKAVFA 492
Db 455 KDILKALILSSALQAGIVWVNYGVVSA--PFSGFKMSGNGRFLGEYGEHEYTEVKTIV 514
QY 493 KL 494
Db 515 KI 516

```

Search completed: June 24, 2003, 10:13:40
Job time : 41.8822 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:42:20 ; Search time 13.7065 seconds
(without alignments)
1062.435 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSH1.PMTVPKILNGLYRQ MSVDALQNYLQKAVPAKIL 495

Scoring table: BLASTM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA:
1. /cgn2_6/ptodata/1/iaa/5A.COMR.pep.*
2. /cgn2_6/ptodata/1/iaa/5B.COMR.pep.*
3. /cgn2_6/ptodata/1/iaa/6A.COMR.pep.*
4. /cgn2_6/ptodata/1/iaa/6B.COMR.pep.*
5. /cgn2_6/ptodata/1/iaa/PTUS.COMR.pep.*
6. /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	1158.5	45.4	521	4	US-09-221-294-2	Sequence 2, Appli
2	982.5	38.5	510	4	US-09-134-001C-4541	Sequence 4541, Ap
3	949.5	37.2	506	4	US-09-134-001C-4383	Sequence 4383, Ap
4	921.5	36.1	485	4	US-09-651-941-9	Sequence 9, Appli
5	921.5	36.1	485	4	US-09-955-547-9	Sequence 9, Appli
6	921.5	36.1	508	4	US-09-655-270A-4	Sequence 2, Appli
7	888	34.8	497	1	US-08-513-841-2	Sequence 2, Appli
8	888	34.8	497	2	US-08-696-834-2	Sequence 2, Appli
9	888	34.8	497	2	US-08-942-674-2	Sequence 2, Appli
10	888	34.8	497	4	US-08-118-317-2	Sequence 2, Appli
11	775.5	30.4	488	4	US-09-144-061C-433F	Sequence 433F, Ap
12	763.5	29.9	493	4	US-09-134-001C-4388	Sequence 4388, Ap
13	685.5	26.9	518	4	US-09-134-001C-4451	Sequence 4451, Ap
14	684	26.8	487	4	US-09-351-224F-5	Sequence 5, Appli
15	657.5	25.8	482	4	US-09-155-183-4	Sequence 5, Appli
16	365	14.3	464	4	US-09-134-001C-4701	Sequence 4701, Ap
17	338.5	13.3	133	1	US-08-346-611-2	Sequence 2, Appli
18	338.5	13.3	133	1	US-08-794-494-2	Sequence 2, Appli
19	194.5	7.6	551	4	US-08-952-661-2	Sequence 2, Appli
20	152	6.0	711	4	US-09-114-001C-5440	Sequence 5440, Ap
21	124.5	4.9	1002	1	US-09-268-317-24	Sequence 24, Appli
22	115.5	4.5	510	1	US-08-489-733-1	Sequence 1, Appli
23	115.5	4.5	510	1	US-08-993-581B-1	Sequence 22, Appli
24	115.5	4.5	510	4	US-09-134-078-22	Sequence 22, Appli
25	108.5	4.3	255	4	US-08-858-207A-465	Sequence 465, App
26	104	4.1	495	4	US-09-134-001C-3128	Sequence 3128, Ap
27	164	4.1	1433	4	US-09-376-330-16	Sequence 16, Appli

28 102 4.0 72 4 US-09-655-270A-33 Sequence 33, Appli
29 102 4.0 509 4 US-09-134-078-18 Sequence 18, Appli
30 100.5 4.4 555 2 US-09-497-027A-6 Sequence 6, Appli
31 99.5 3.9 2314 4 US-09-268-347-49 Sequence 49, Appli
32 98 3.8 687 1 US-08-164-839-31 Sequence 31, Appli
33 98 3.8 587 1 US-08-164-839-31 Sequence 33, Appli
34 98 3.8 687 1 US-08-583-799-31 Sequence 31, Appli
35 98 3.8 587 1 US-08-583-799-33 Sequence 33, Appli
36 98 3.8 588 1 US-08-104-839-70 Sequence 70, Appli
37 98 3.8 588 1 US-08-164-839-72 Sequence 72, Appli
38 98 3.8 688 1 US-08-583-799-70 Sequence 70, Appli
39 98 3.8 688 1 US-08-583-799-72 Sequence 72, Appli
40 98 3.8 2404 4 US-09-134-001C-3464 Sequence 3464, Ap
41 97.5 3.8 454 4 US-08-984-618-13 Sequence 13, Appli
42 97.5 3.8 495 4 US-09-217-490-2 Sequence 2, Appli
43 97.5 3.8 662 1 US-08-224-657-94 Sequence 94, Appli
44 97.5 3.8 662 4 US-09-354-138-94 Sequence 94, Appli
45 97.5 3.8 1053 4 US-09-724-510-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-221-294-2
; Sequence 2, Application US/09221294
; Patent No. 6268138
; GENERAL INFORMATION:
; APPLICANT: Riccardo Dalla-Favera and
; APPLICANT: Alessandro Massimo Gianni
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: Adenoviral Vector Capable of Transducing the
; TITLE OF INVENTION: Vector
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER PREPARED FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09221,294
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/INVENTION NUMBER: 4,940-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-291-0525
; INFORMATION FOR SEQUENCE CHARACTERISTICS:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-221-294-2

Query Match 45.4% Score 1158.5 DB 4 Length 521
Best Local Similarity 48.8% Pred No 8 20-100
Matches 235, Conservative 88, Mismatches 152, Indels 7, Gaps 5;
QY 20 QPTGLFINNKVPKQKTEVINPSTEEETICHIEEDVEAVQAAADPAFSGS-WN 78
DQ 15 QYTKIFINNEHDSVSKKFTVFNFATEEELQVEEDCKEDV-KAVKAAQAFQETGTFW 94
QY 79 GIDIDPDKALYPALAEILIEQK--IVIASIEILNCKAISSS-PCIVDIDVINVKSSAGF 135


```

Matches 189; Conservative 90; Mismatches 191; Indels 7; Gaps 3;
QY 24 LFINKKVPKONKTEVNPSTEEBICHLYGREDDEVEAOADRAFSNCSWNGIDPI 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 33 LVIGDQITPSTGAFUSINPAQSHLASVAEATAAUAFAVFAAKAAAR--TWQMPFA 90
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 DRGKALYLAFLPQDDVDTASTFTLNGKKAISSEPG-IVDLVINYLKSSAGPADKIDCP 142
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 91 QRTKLMPKYAALIEHKTELAQLOSKDMGKPIRESUGIDLPIMETLEYFAGLVTKIEGR 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 MIDRPTHPSTYKRPQIGVGGQITIPWNPFLMMAWKIAIPALVTCNTVVLKTAESTPLSAL 202
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 151 TTPAPGRFLNLTREPIGVGVAITPNPFAVOAVKVIAPALAMGNAIVLKPQAAPLVVPV 210
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 YVSKYIPACGIPGCVINIVSGFKIVVEALTNHPKIKKVAFTGCTATCCHITVOSAAAGLK 262
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 211 ALGHALEAGLPGLVNLVPGRSVAGNALVQHPSVKVTFTGSTEVGQOIGKMAADRLLI 270
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 KVTLELGKSPNIVFADAELKKAQVNIILGIYVNSGVCAGSRVYVEESYDKFIEEFK 322
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 TASLELGKSAVAFGDSKPAVAAVVQAMYSNQGETCTAPSRLLVERPIDEVVELVQ 330
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 323 AASESIKVGPDFDETFQGAQTQOMQLNKILKYVDIGKNEGATLTGGER----LGSKY 378
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 ARVEAARVGGPDPDPTETIGPLTSAEQGRESVHSVWSNTEEGATLISGQPSPTGAPEQGF 400
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 FIKPTVFGDKEDMRIVKREIFGVPVTVTKFSADDEVINMANUSEYGLAAGIHTSNINTA 438
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 391 YYRTLESGVTADMTAREIFPVLISVLPFGFEBAITLANVTFGLAAGVETPDVQWA 450
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 439 LKVAADRVNAGTVINTYDHPHVAPEGGNAGLGRMSVDALQNYLOVKAVRAKLD 495
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 451 LRFAGTLDAGNVNWSGVLNPASPYRGQSGYSGSDIQAATIESPTKEKSIWAPLD 507
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-513-841-2
; Sequence 2, Application US/08513841
; Patent No. 5753481
; GENERAL INFORMATION:
; APPLICANT: Niwa, Minco
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5753481el L-sorbose dehydrogenase and No. 5753481el L-sorb
; TITLE OF INVENTION: dehydrogenase Obtained from Gluconobacter oxydans T-100
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/513,841
; FILING DATE: 01-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9304700.9
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 241851/1993
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. ORLON
; REGISTRATION NUMBER: 24,618

REFERENCE/LOCKET NUMBER: 18-909 0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704-413-3000
TELEFAX: 704-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Gluconobacter oxydans
STRAIN: T-100
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..497
IDENTIFICATION METHOD: experimentally
US-08-513-841-2

Query Match 34.88; Score 888; DB 1; Length 497;
Best Local Similarity 39.88; Pred. No. 1.5e-74;
Matches 194; Conservative 93; Mismatches 189; Indels 12; Gaps 5;
QY 6 MIVPEKLPNLEVEGQGTGFTLNKPFVSKUNKFEFVINSFEFTFCHIVFDETFVEAV 65
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 VSLPIK-----PREGFTLKEWKACKLI- FEDESSFAHDVVPVTRIPRTRELDKAV 57
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 GAALRAHSNWSHPTDHPGKALYKIALGFLDQKRIASIELLNGKALSSSGIDVLY 125
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 AAARRAFENGSWAGLAAADRAAVLLKAAAGLLRRERDIAVWELENKPIISQAKGEIDH 117
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 INVLKSSAGFADKIDGPMIIT- GPTHFSVTKGQPLQVWGLTPWNPFLIMMAWKIAPALV 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 IACFEMAAAGARMHGITENNLAGELGGLGVLEKPIGVVGLIPWNPFLIMLTERAPFLA 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 PNTVVLKTAESTPLSALYVSKYTFQATIDPQVINIVSGP-KIVVEALTNHFKKVAFT 244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 SGTFLVVKPAEVTSAITLLAFLLALAGLAKVENNVITGKRI-VQAMTEHQDILMSEPT 247
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 GSTATGRH-IYUSAAAGLKKVILELAKSKSNIVFAVLAELKKAQVNIILGIYVNSGVC 404
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 GSTGVGKSCIHAAANSNKKLELEGENE-VVIALSRLELAATAVAATVLENDQGVVS 297
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 GSPVYVEESIYDKFIEEFKAASES IKVGDPEFESTFGCAQTSOMQINKILKYVDIGKNEG 463
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 298 SSPLVPSVAPKPPFVVKPKKIPVGPDPPTFQICATITTEAUNKTILYIANGKARG 457
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 ATLTGGERLG-SKGYFIKPTVPGVKNKEMPVREIFTPVVTVTFKSAIEVINNAEIS 422
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 458 AKLGGGTVIDGKQGVQPTFTFVKKPSMLIARDIEFGVGLASFEHFDIVFAIAIANH 417
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 EYGLAAGIHTSNINTALKVADRVNAITVWNTINYDHPHVAPEGGNAGLGRMSVDALQ 482
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 418 VYGLAASVNSKIDKALAVTRAKVA-RFVWNIMSGGFTLPLAGKSSWVKEA-LYVVE 477
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 483 NYLQVRKV 490
| : : : :
DB 478 EYTOIKSV 485
| : : : :

RESULT 8
US-08-696-834-2
; Sequence 2, Application US/08696834
; Patent No. 5834263
; GENERAL INFORMATION:
; APPLICANT: Niwa, Minco
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Hayashi, Hiromi
; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonate Acid
; NUMBER OF SEQUENCES: 48

```


Best Local Similarity 39.88; Pred. No. 1.5e-74;
Matches 194; Conservative 93; Mismatches 189; Indels 12; Gaps 5;
QY 6 MTPVPIKPLNGLEYEQPTGLFNNKVPFSKONKTFEINPSTEEIEICHIVGREGDEVEAV 65
DB 7 VSLPLK-----PREGFFIDGEMWAGKD--FFDHSPPAHVPVTRIPRCTPEIDLEAV 57
QY 66 QAAKRAFSNSWNGIDPUDGSKALYHIAFLHUKIVIASIELUNGKAISSSPGVDLV 125
DB 58 AAARRAFENGSWAGLAADRAAALLKAAAGLLRRRDDIAYWEVLENGKPISOAKGEIDHC 117
QY 126 INYLKSSAGPADKIDGFMIDT--GRTFESYTKFQPGVCGUQIIPWNEPILLMMAWKIAPALV 184
DB 118 IACFEMAAGAARMHLGDTFNNLCEGLFGWVLEPVGVLITPWNFPFIMLCERAPPIIA 177
QY 185 TGNVTVLKTAESTPLSALYVSKYIPGAGIPPGVINIVSGFGLVVEATINHPKIKKVAFT 244
DB 178 SGCTLVVKKPAEVSATILLAEILLALAGLPGKVFVNVVGTGTGTGQAMTEHODIMLSFT 237
QY 245 GSTATGRH--IYQSAAGLKKVTLLEGGKSPNIVFADAEIKKAVQNIILGLIYNSGFVCA 303
DB 238 GSTGVGSCTHAAASNLKGLJELGGKNIIVFALSNLEDAADAVAFGISENIGQCVS 297
QY 304 GSRVIVLESYDKELEFKAASESIVKGDPEDETFQGAQISOMQLNKILKYVDIDGKNG 363
DB 298 SSRILIVERSAERFERLVVPMKIRVGDPEDETFQGAQISOMQLNKILKYVDIDGKNG 357
QY 364 ATLTGGERLG--SKGVFIKPTVGDYKEDMRIVKEIFGPVVTIVTKFSADSEVINMADS 422
DB 358 AKILCGGIVDQKGYIQPTLTVVKSMLARDEIFGPVLAHFDTVDALAIANDT 417
QY 423 EYCLAAGIHTSNITALKVADRVNAGTVWINTYDFHHAVPFGFNASGLGREGMSVDAIQ 482
DB 418 VYGLAASVWSKIDKALAVTRVRAGRFVWNTIMSGPETPLGGFKQSWGREGALYGV 477
QY 483 NYLQKAV 490
DB 478 EYTOIKSV 485
RESULT 10
US-09-118-317-2
; Sequence 2, Application US/09118317
; Patent No. 6197562
; GENERAL INFORMATION
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 6197562-1 L-sorb-se L-hydrogenase and No. 6197562-2
; TITLE OF INVENTION: L-sorbosone behydrogenase Obtained from Gluconobacter
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS
; ADDRESS: Obolon, Spivak, McClelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS-DOS Editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,317
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/514,841
; FILING DATE: 01-NOV-1995

APPLICATION NUMBER: US 9404700.9
FILING DATE: 08-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 241851/1993
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-909 0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-4000
TELEFAX: 703-413-2220
FLEX: 248855 OPAT DR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORGANISM: Gluconobacter oxydans
STRAIN: T-100
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..497
IDENTIFICATION METHOD: experimentally
US-09-118-317-2
Query Match 34.8%; Score 888; DB 4; Length 497;
Best local similarity 39.88; Pred. No. 1.5e-74;
Matches 194; Conservative 93; Mismatches 189; Indels 12; Gaps 5;
QY 6 MTPVPIKPLNGLEYEQPTGLFNNKVPFSKONKTFEINPSTEEIEICHIVGREGDEVEAV 65
DB 7 VSLPLK-----PREGFFIDGEMWAGKD--FFDHSPPAHVPVTRIPRCTPEIDLEAV 57
QY 66 QAAKRAFSNSWNGIDPUDGSKALYHIAFLHUKIVIASIELUNGKAISSSPGVDLV 125
DB 58 AAARRAFENGSWAGLAADRAAALLKAAAGLLRRRDDIAYWEVLENGKPISOAKGEIDHC 117
QY 126 INYLKSSAGPADKIDGFMIDT--GRTFESYTKFQPGVCGUQIIPWNEPILLMMAWKIAPALV 184
DB 118 IACFEMAAGAARMHLGDTFNNLCEGLFGWVLEPVGVLITPWNFPFIMLCERAPPIIA 177
QY 185 TGNVTVLKTAESTPLSALYVSKYIPGAGIPPGVINIVSGFGLVVEATINHPKIKKVAFT 244
DB 178 SGCTLVVKKPAEVSATILLAEILLALAGLPGKVFVNVVGTGTGTGQAMTEHODIMLSFT 237
QY 245 GSTATGRH--IYQSAAGLKKVTLLEGGKSPNIVFADAEIKKAVQNIILGLIYNSGFVCA 404
DB 238 GSTGVGSCTHAAASNLKGLJELGGKNIIVFALSNLEDAADAVAFGISENIGQCVS 297
QY 304 GSRVIVLESYDKELEFKAASESIVKGDPEDETFQGAQISOMQLNKILKYVDIDGKNG 463
DB 298 SSRILIVERSAERFERLVVPMKIRVGDPEDETFQGAQISOMQLNKILKYVDIDGKNG 457
QY 364 ATLTGGERLG--SKGVFIKPTVGDYKEDMRIVKEIFGPVVTIVTKFSADSEVINMADS 422
DB 358 AKILCGGIVDQKGYIQPTLTVVKSMLARDEIFGPVLAHFDTVDALAIANDT 417
QY 423 EYCLAAGIHTSNITALKVADRVNAGTVWINTYDFHHAVPFGFNASGLGREGMSVDAIQ 482
DB 418 VYGLAASVWSKIDKALAVTRVRAGRFVWNTIMSGPETPLGGFKQSWGREGALYGV 477
QY 483 NYLQKAV 490
DB 478 EYTOIKSV 485
RESULT 11
US-09-134-001c-4246
; Sequence 4246, Application US/09/14001C
; Patent No. 6380370


```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4246
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4246

Query Match      30 4%, Score 775.5; DB 4; Length 488,
Best Local Similarity 36.6%; Pred. No. 56-64;
Matches 174; Conservative 97; Mismatches 188; Indels 17; Gaps 7;

QY 22 TLEFNKVFVKNTFEVINFSTEEIECHYGFPGDVEEAVQAALRAFSNGSNWILH 41
DB 18 TKQINGEWDVSASGETIDVINPATIEVMGKIAGNEEDVNKAADAKVYLFFRHSVE 77
QY 82 PIDFCKALYLAELIEQDKVIASTFTLEWFKATSSSPGIVIVNYL...LKSSAGFAUK 138
DB 78 --ERRELLKIVKEYQKKNKDLIPATITTEIGAVLSVEN-----VHYQMELNHTAARDA 130
QY 139 INGRMIDGTRHFSYKPEQIPVGGIIPWNEFLIMAWKIAIAPVTCNTVVLKTPAFSP 198
DB 131 LDSQFEBQKGU-DLVKKEALGVASLVTPWNEFLINISLALAAFAAGSHVVLKPSSETP 189
QY 199 LSALYVKYIFQAGLIPGVINIVSGFGKIVVEAITNHPKIKKVAFTGSTATGRHIYQSA 258
DB 190 FAAILILAFIDFKVGPVGVNLVNGDSGVGNPLSEHPKVRMMSFTGSGTKIMEKAA 249
QY 259 AGLKVKVLELGKSPNIVFADELKKAQVNIILGTYVNSGEVCCAGSRVYVESIYDKFI 318
DB 250 KDFKVSLELGKSPYIVLDVVDVEEAANAIIKKVNNNTQVCTAGTRVLIPESIKEDYL 309
QY 319 EEFKAASIESKVGDFPESTFGAGTSTQMUNILKLYVDIGKNEGATILIGG----ERLG 474
DB 310 TAVKRAFSKVRVGPQREKGTQVGPILSKKQFDQVQYIDKGINSTAEILPVGSGKPKEDL 468
QY 375 SKGYEIKPTVGDVKEDMRIVKPEIFGVVTVTKFKSADRVINMANSYGLAAGIHTSN 434
DB 366 UKGYFAPPTIFINLNMHTIAQREIFHPVMSVITYNNLDAEIAIANUTYGLAGYVIGKD 428
QY 435 INTALKVADRYNAGTIVNITVNDHFHVPFGGFNASGLGREGMSVIALNYLAKVAV 490
DB 429 KDTLPHVARSIEAGTIEINEAGR-KPELPFGGKESGLGPEFWGYGIEEFLEVKS 483

RESULT 12
US-09-134-001C-4388
; Sequence 4388, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4388
; LENGTH: 493

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4388

Query Match      29.9%, Score 763.5; DB 4; Length 493;
Best Local Similarity 34.5%; Pred. No. 6,80-63;
Matches 162; Conservative 109; Mismatches 181; Indels 17; Gaps 6;

QY 24 LFINKKVFVKNTFEVINFSTEEIECHYGFPGDVEEAVQAALRAFSNGSNWILH 83
DB 24 LFINKKVFVKNTFEVINFSTEEIECHYGFPGDVEEAVQAALRAFSNGSNWILH 81
QY 84 DPGKALYLAELIEQDKVIASTFTLEWFKATSSSPGIVIVNYL...KSSAGFAUK 141
DB 82 TRACHVKLLIPLEKKNKDELAOLYVKEGKILAGAGELIDKISFIDYMTLSMSUK--G 139
QY 142 PMIDTGFPHFS-YTKPQIPVGGIIPWNEFLIMAWKIAIAPVTCNTVVLKTPAFSP 200
DB 140 PVLGNSTANETIQIINKPIGVTAGIVPWNAPILVLMKVPVPAIVTGSVVVVKPSSEPT 199
QY 201 ALYVKYIFQAGLIPGVINIVSGFGKIVVEAITNHPKIKKVAFTGSTATGRHIYQSA 250
DB 200 TIFLAELIPASTIPALVIVGSGFTVTCGASHKELFESTIGSMFAKESVYPAVAT 259
QY 261 IKVYTIELGKSPNIVFADELKKAQVNIILGTYVNSGEVCCAGSRVYVESIYDFIEE 320
DB 250 VKKNVLELGNAPVIVISNALFLKAVNYIVTAPINSAGVCTGHEFTVIEVWHDGFINP 319
QY 341 PKAASESKVGDFPESTFGAGTSTQMUNILKLYVDIGKNEGATILIGG----ERLG 480
DB 340 VTSKMSLSLVGDPEDENTVGCATINQKQDSIHEKVQDAIFKNCATIMTGHQIKPHCFEY 474
QY 381 KPTVGDVKEDMRIVKPEIFGVVTVTKFKSADRVINMANSYGLAAGIHTSN 440
DB 380 APTVLIDNRKDYVNFKDEIFGVLAIITYRDFEQTIEDANDTNGLUSSYIFSSELEVTMT 439
QY 441 VADRYNAGTIVNITVNDHFHVPFGGFNASGLGREGMSVIALNYLAKVAV 484
DB 440 ATERLKPGEVYANCEAEVWNGYH----AGWRESGLGGADGTHGFEY 483

RESULT 13
US-09-134-001C-4451
; Sequence 4451, Application US/09134001C
; Patent No. 6480470
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4451

Query Match      26.9%, Score 585.5; DB 4; Length 518;
Best Local Similarity 33.1%; Pred. No. 1,5e-55;
Matches 173; Conservative 90; Mismatches 200; Indels 51; Gaps 14;

QY 6 MTPVPIKPLNGLEYEQPTG-----LFINKKVFVKNTFEVINP- 44
DB 5 MWVPEKNEPGLDFSVQTNVEREFNEELKVKRAQLQJQIPLVINGEKL--TKIDTFNSNVA 62
QY 45 STEEIECHYGFPGDVEEAVQAALRAFSNGSNWIDPDKALYLAELIEQDKVIA 104
DB 45 STEEIECHYGFPGDVEEAVQAALRAFSNGSNWIDPDKALYLAELIEQDKVIA 104

```

63	NTS:I:IAKV:ATQDDIFKAF:ESANHA:Y---SWKKWSHKDRAE:ILL:EVAA:II:PRPK:EE:IS	120
105	STETID:NGKAI:SSSRGD---VDIAVINYLKSSAGFADKIDGR-MIDTGRTHSYTKRQPL	159
121	ALIMVEAGKPDEAVGDAAG:IDPFI:YYA:SRMELA---DGRKPIDREG:EHNP:YIK-PI	176
160	GVCCQI:IPWNP:PLIM:AWK:IPALV:IGNTV:VLK:IAES:IPLS:ALYS:VSKY:IPQAG:IPGV:IN	219
177	GTGVT:IPWNP:PF:AIMAG:TTI:PAVAG:NTV:LLK:PAED:TVILTAYKLM:ELIEEAGL:PGV:VN	236
220	IVSGFGKIVVEA:INHPKIKK:VAF:TSAT:GRH:IIYOSAA-----GLAKKV:LELGGKSP	273
237	FVPGDPEIGDY:IVD:HKDTH:VFTGSRATG:TKIYERSAVV:GGQOFLKRV:IAEMGGKDA	296
274	NIVFADAE:ELKAVQNI:ILGIY:INSGEVCCAGSRVYVEESIYDKFIEEFKAA:SEIKVGD	333
297	IWDNNVDILAAE:AI:VTSAF:GSGQKCSAC:SR:AI:VHQV:VHDE:ILEKAIQLKULIGNI	356
334	FDESTFOCAOT:SQMOLANKILKYVDI:CKNECATILITGGERLGSKCYEIKTVF:GCDVKEDMR	393
357	-EENTFMG:VINOKOFIKINYIEIGKKEG-KLETGGGTDDSGYIEPTIFSG:QSADR	414
394	IVKEEIPGVVTVTKPKSAD:EPINMAND:SEGLAAGHTISNTALK:KADVRNAGT:VWIN	453
415	IMQEEIPGVVGFVKVDF:DEA:LEVANDTDYGLTGAVITNHR:EHMIAKYNFEFVGNLYIN	474
454	-----TYNDPHIHAVPF:GFGFNAG:IL-GREMSVDALQNYLQVAK	490
475	RGCTAAVVGYH---PFGGFKMSGT:TDARTGSPDYLLNFELEKVV	514

```

RESULT 14
US-09-351-224E-5
: Sequence 5, Application US/09351224E
: Patent No. 6388171
: GENERAL INFORMATION:
: APPLICANT: Duvick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob
: APPLICANT: Folkerts, Otto
: APPLICANT: Crasta, Oswald P.
: TITLE OF INVENTION: Compositions and Methods for Fumonisin
: DETOXIFICATION
: FILE REFERENCE: 5718-111
: CURRENT APPLICATION NUMBER: US/09/351,224E
: CURRENT FILING DATE: 1999-07-12
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 487
: TYPE: PRT
: ORGANISM: Exophiala spinifera
US-09-351-224E-5

```

	Query Match.	26.8%	Score	684;	DH 4,	Length	487;
	Best Local Similarity	33.4%, 94	Mismatches	184; Indels	36;	Gaps	12;
	Matches	164;	Conservative				
QY	19 EQPTGLFINNKFFVPKSKUNTKFEVINPSTBEETIC-HIYEGRDDVEEAVQAADRAFSNCSW	77	:	: : :	:	:	:
DB	7 EYKSELPINNEFVSSGSGSEKLITNPWDSTVAIDVVHVANAADVSAVASAQVKKGKP	66	:	: : :	:	:	:
QY	78 NGDPIIRGKALRYLAPLIIPQNKDVIASFTTLGNKAISS-SPEVDVLIVNYLKSSAGFA	136	:	: : :	:	:	:
DB	67 KKFTGAQRACMLKFAPLAEKNAEKLEJESLTGPSPSMITHFDIPNMVSFRYYAWA	126	:	: : :	:	:	:
QY	137 DKIDGRMI---DTGRTHPSYTYRKPGVCVGCIILPWNEPLIMAAWK LAPALVTGTNTVYLTA	194	:	: : :	:	:	:
DB	127 DKTAGTTFPEDNGKNRWY---EPMGCVAGIASWNATFLYGCKWTAPALAACCSTIFRAS	183	:	: : :	:	:	:
QY	195 ESTPLSALYSKYTIPOAGTPGGVINIVSGFGXIVVEAITNHPKIKKVAFQTGSTATGRHIU	254	:	: : :	:	:	:
DB	184 EKSLPVIGIALPFAGAGEPPGVQPRTTG-APVTTGEALASHMNIAKISPTRSVSGGPAVK	242	:	: : :	:	:	:

```

255 QY USA--AAGLKKVYTLGGKSGSPNIVFAUAEKKAVUNILILYVNSGVPVYAGSK VYVE 411
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
243 Db QATLKSNMKPVLEILGEE-PTIVNEAPLER 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 311 ESI-----YDKFEEKKAASESIKVG DPDESTHQAGTSGMGI NKILKYVDIG 459
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 241 ESYLVVWNLAFKPEVPP-SFQVLPWFQGNP EKKPDDGVVIFQUTDRVLENTDVG 490
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 360 KNEGAAILLIGGERAGSGGYEIKPTGVGVKREIMRIVKEEIPGVVTVTKESADEVINMA 419
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 351 K-GTAQLLIGVSPKGNKEALEPIFVNPKPGSKWPEEIPGVISIKTEKEEATETA 409
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 420 NUSEVGLAAGTHTSNITALKVAORVNAVIGVWINTYDHFHIAVPPGCFPNASGLGKPMVD 479
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 410 NGTGVGLASVIYTKSLNKGTVSSALEFGVSIINFEPTFQVETKMKQSGSGELGEE 469
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 480 ALONYIQVKAV 490
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 470 GLKAYLEPKTI 480
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 15
US-09-155-183-4
: Sequence 4, Application US/09155183
: Patent No. 6424011
: GENERAL INFORMATION:
: APPLICANT: Narbad, Arjun
: APPLICANT: Rhodes, Michael J.C.
: APPLICANT: Gasson, Michael J.
: APPLICANT: Walton, Nicholas J.
: TITLE OF INVENTION: PROPERTIES OF VANILLIN
: FILE REFERENCE: 20747/100
: CURRENT APPLICATION NUMBER: US/09/155,183
: CURRENT FILING DATE: 1999-05-03
: EARLIER APPLICATION NUMBER: PCT/GB97/00809
: EARLIER FILING DATE: 1997-03-24
: EARLIER APPLICATION NUMBER: GB96/06187
: EARLIER FILING DATE: 1996-03-24
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 482
: TYPE: PRT
: ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

Query Match	25.8%	Score	457.5	DB	4	Location	4827
Best Local Similarity	32.9%	Prod. No.	5,40-53				
Matches	156	Conservative	90	Mismatches	205	Indels	19
QY	24	LFINKKVPKONKTFIVINUSITFEETFTCHLYTFRERHVFEEVAGVAAARAFESNSNNHLLPT	83				
DB	6	LLGGQSPADMGKTFERRNIVGTLLYSRVAAATLEDAVAAVAAQQAHP	63				
QY	84	TPKALVPLAFLEKQVIVLASITELLINIKATISSRQIVGVIVINIKSSAIFAKITP	142				
DB	64	ERRSKLLKAAFLQARSGEEVTEADGETGAMANNVGENVLAANNMLREAAASMTLVGNCE	121				
QY	143	MDTG-ETHESYTKRQPLQGVQVITPNEFLMHWAKTAPALVIVNTVVLTKESIPISA	201				
DB	122	VPSDVPGPSAMALRQGVGVLLIATWNAHVITATKATIMTAVNTVVKASEIPAVH	181				
QY	202	LYSKYTFQVATFVIVIVSYQVRAVLAIVNRPKFRFVATCTAPPHLYGAA	298				
DB	182	RLCGVLQDAGLGVGVNVISNAPADAAQVLEKLTANIVARRVNTFGSTHGVIRVITSLA	241				
QY	259	AGLEKKVILEGGSPNIVFADAEELKAVQNTILGITYNSFGVGVWNSPVVDESYVERET	318				
DB	242	RHLKALLEGGKAPLVLIDDALEAAVAAVAAFGAYFNQGVQIMSTERIVIAKVAIAFV	301				
QY	319	EEFKAASESTKVGIPPESTFEAGVTSQMLAKTLKLVIGKNEGATILLEGEGEGVY	378				

```

Db      302 AQLAAKVETLRAGDPADEFSLVSSILVDSACTRIKALIDUAVAKGAVLVIGGQLEGS--- 358
QY      379 FIKPTVEGDVKREDMFIYKEEIFGPPVTVTKREKSADDEVINMANDSYGLAAGIHTSNINTA 438
Db      359 IIOPTLLDVTASMPFYREESFGPVAVVLEFGGFEALIQIANDSFFGLSAAIFSPDTGPA 418
QY      439 LKXADRVNAGTWIN--TYNDFHHAVPPGGGFNASGLGREMSVDALQNYLOVKAV 490
Db      419 LALAORVESGICHINGPTVHD-EAOMFPGGVKSSGGYSGFGGKASIEHFTOLRWV 471

```

Search completed: June 24, 2003, 10:31:38
 Job time : 15.7085 secs



Genome version 5.1.6
Copyright (c) 1993-2003 Cambridge

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 : Search time 23.9275 Seconds
(without alignments)
2238.528 Million cell updates/sec

Title: US-09-830-751-2
Sequence: 1 MSHUPMTVPIKUPNILEYEL MSVDAIQNYICVKAIPAKILD 495

Scoring table: MS-09-830-751-2
Gapop 10 0, Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database: Published Applications-AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUB pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUB pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_PUB pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUB pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUB pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUB pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US06_PUB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1399	54.9	496	US-09-847-208-59	Sequence 59, Appl
2	1362.5	53.4	495	US-09-847-208-11	Sequence 11, Appl
3	1218.5	47.8	512	US-09-961-403-12	Sequence 12, Appl
4	1218.5	47.8	512	US-10-268-518-2	Sequence 2, Appl
5	1218.5	47.8	512	US-10-305-823-14	Sequence 14, Appl
6	1201	47.1	538	US-09-344-882-20	Sequence 20, Appl
7	1201	47.1	538	US-10-293-865-20	Sequence 20, Appl
8	1169.5	45.9	534	US-09-344-882-24	Sequence 24, Appl
9	1169.5	45.9	534	US-10-293-865-24	Sequence 24, Appl
10	1146.5	45.0	501	US-09-444-882-22	Sequence 22, Appl
11	1146.5	45.0	501	US-10-268-518-4	Sequence 4, Appl
12	1067	41.8	492	US-10-175-696-21	Sequence 21, Appl
13	1056.5	41.4	493	US-09-823-901-9	Sequence 9, Appl
14	1056.5	41.4	493	US-09-815-242-10550	Sequence 10550, A
15	1025	40.2	496	US-08-781-986A-5241	Sequence 5241, Ap
16	1015	39.8	508	US-10-166-087-4	Sequence 4, Appl
17	984	38.1	500	US-09-815-242-5644	Sequence 5644, Ap
18	970.5	38.1	496	US-09-815-242-15657	Sequence 15657, A
19	970.5	38.1	496	US-09-815-242-15657	Sequence 15657, A

20	948	34.8	490	US-09-815-242-1262	Sequence 1262, A
21	929	36.0	490	US-09-815-242-10657	Sequence 10657, A
22	899	40.1	505	US-09-738-626-6572	Sequence 6572, Ap
23	878	34.5	412	US-09-925-302-736	Sequence 736, App
24	878	34.4	518	US-09-919-039-143	Sequence 143, App
25	845.5	32.8	487	US-10-175-696-17	Sequence 17, Appl
26	845.5	32.8	487	US-09-823-901-5	Sequence 5, Appl
27	828.5	32.5	414	US-09-925-302-148	Sequence 148, Ap
28	828.5	32.4	481	US-09-815-242-13829	Sequence 13829, A
29	781.5	30.6	475	US-09-815-242-12475	Sequence 12475, A
30	764.5	30.0	482	US-09-815-242-18264	Sequence 18264, A
31	764.5	30.0	497	US-09-748-626-4108	Sequence 4108, Ap
32	761.5	29.9	483	US-09-815-242-6584	Sequence 6584, Ap
33	754.5	29.6	482	US-09-815-242-14047	Sequence 14047, A
34	735	28.8	490	US-09-748-626-3558	Sequence 3558, Ap
35	586	25.3	481	US-09-748-626-6440	Sequence 6440, Ap
36	584	25.8	487	US-09-882-994-5	Sequence 5, Appl
37	562	26.0	491	US-09-947-684-3	Sequence 3, Appl
38	557.5	25.8	482	US-09-733-383-4	Sequence 4, Appl
39	551.5	25.5	498	US-10-272-419-12	Sequence 12, Appl
40	637	25.0	481	US-08-976-0630-28	Sequence 28, Appl
41	622	24.4	521	US-09-738-626-6482	Sequence 6482, Ap
42	591.5	23.2	439	US-09-815-242-13316	Sequence 13316, A
43	584	24.1	483	US-09-748-626-4047	Sequence 4047, Ap
44	579	22.7	584	US-09-748-626-3680	Sequence 3680, Ap
45	472	18.5	882	US-09-838-561-2	Sequence 2, Appl

ALIGNMENTS

FFSULT 1
US-09-847-208-59
: Sequence 59, Application US/09/847208
: Publication NO: US20030082150A1
: GENERAL INFORMATION:
: APPLICANT: Saxon, Andrew
: APPLICANT: Zhu, Daoheng
: TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
: TITLE OF INVENTION: IGR-MEDIATED ALLERGIC DISEASES
: FILE REFERENCE: CC67-003A
: CURRENT APPLICATION NUMBER: US/09/847,208
: CURRENT FILING DATE: 2001-05-01
: NUMBER OF SEQ ID NOS: 177
: SOFTWARE: FastSPQ for Windows Version 4.0
: SEQ ID NO 59
: TYPE: PRT
: ORGANISM: Cladosporium herbarum
US-09-847-208-59

Query Match	54.9%	Score 1399	PR 9	Length 456
Best Local Similarity	53.8%	Prod No 4	4e-98	
Matches	263	Conservative	80	Mismatches 136
				Indels 2
				Gaps 2
Q7	7	IVLLEFNGLEYDFTGLFNNKVFVS	KONKTFEIVINFEETEEICHIVREEDW	VERAVQ 66
Db	3	SVOLETHISKYEQFTGLFNNKVFVS	KONKTFEIVINFEETEEICHIVREEDW	VERAVQ 62
Q7	67	AAHAFNSNSWNSDPTDCKALYHIAF	LECKKNVASIFLNDKATSSSPGLDLV	126
DE	63	AARQAFEGSWRETEFENRGLNNLAN	FLKFNKTLAAVESIDNGKATSMARVTS	121
Q7	127	NYLKSSACFARKINGMRTGTFHESV	TRKPDQGVGGQIINWNEFLMMAWIA	186
DE	122	QGLPYFYSWALKITGKVLGISTPT	PHVYKPEKATVPSQUSPFLHMAW	181
Q7	187	NTVVKRTAFSPISALYVSKYIDNAC	ITPCIVINVSFQKIVVEALINHK	181
DE	182	NTVVKRTAFSPISALYVSKYIDNAC	ITPCIVINVSFQKIVVEALINHK	181
Q7	247	TATCPHYQSSAAAGLKKVTLFLD	GRSPNIVAFATKRAVQNIHLS	181

```

Db 242 TVGRTILKAAASNLKRVTLLEGGKSPNIVFEDADIDNAISWVNFIFNHHGCCAGS 301
Qy 306 RVVVEFSYDKFIEEFKAASESIKVGDPDEFDESTFQGAQTSOMQLNKLKLYVDIGNEGAT 365
Db 302 RVVVGESYDKFKVKEKRAKVVVGUPFAADIFQ3QVSKVQFQFIMFYIQAQKIAGAT 361
Qy 366 LITGGERIKSGGYFIPKTPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEYG 425
Db 362 VETGSRKSGDKGYFIETPFSNVTEDMKIVKEEIFQPVGSIAKFKTKDAIKLGNASTYG 421
Qy 426 LAAGIHTSNINTALKVADRVNACTVWINTYNDFHHAVPFGGPNASGLGKREMSVVALONYL 484
Db 422 LAAAVHTKNINTALEVSNALKAGTVWVNTYNTLHHOMFEGGKESGIGRELGEDALANT 481
Qy 486 QVKAARAKL 494
Db 482 QIKIVSIRL 490

RESULT 2
US-09-847-208-11
; Sequence 11, Application US/09847208
; Publication No. US20040082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxen, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daoheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IgE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: US67,002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Alternaria alternata
US-09-847-208-11

```

```

Query Match 53.4%; Score 1362.5; DB 9; Length 495;
Best Local Similarity 54.3%; Pred. No. 2,66-95;
Matches 266; Conservative 84; Mismatches 135; Indels 5; Gaps 5;

Qy 7 TVPKLPLNGLEYDPTGLFINNKVFPSKONKTFEVINISTEEELICHYEGREDVEFAVQ 66
Db 3 SVKLSTPQTGFFQPTGIFINNEFKVAVGKTFDWINISTEEVICSVOEATEKDVDAVA 62
Qy 67 AADAFNSNSWNGIDPTPKKALYELAELEQIKIVIASIETLUNGKAISSRG-DVULV 125
Db 63 AARKAP-NPWAKEPTEPNRGLINKLADLFEKNADIIAAVEALDNGKAFSNKAKVDVPA 121
Qy 126 INYKSSAGFAUKIDCKMIDTSKTHSYTKKQPLAVQGIIPWNEPILMLWAKIAPALVT 185
Db 122 AGPLPYGGWAIKIFKVVVITADPSNYI-FKSLIVFAVRSSMELPILMWKSGICPATAT 180
Qy 186 GNTVLKLTARSTPLSALYVSKYITPQAGIPPIVNIIVSGRGIIVFAFATINHEKIKKVAFTG 245
Db 181 GNTVLKLTARSTPLSALYVSKYITPQAGIPPIVNIIVSGRGIIVFAFATINHEKIKKVAFTG 245
Qy 246 STATGRIHYOSAA-AGLKVVTELEGGKSNIVFAIAELKKAQVNTILGIIYVNSGRVYAG 304
Db 241 STVVGRIKMSAAGSNLKVTELEGGKSNIVFAIAELKKAQVNTILGIIYVNSGRVYAG 304
Qy 305 SRVYVEFSIVDKFIEREKAASESIKVGDPDEFDESTFQGAQTSOMQLNKLKLYVDIGKNEGA 364
Db 301 SRVYVEFSIVDKFIEREKAASESIKVGDPDEFDESTFQGAQTSOMQLNKLKLYVDIGKNEGA 364
Qy 365 TLITGGERIKSGGYFIPKTPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEY 424
Db 360 TLITGGERIKSGGYFIPKTPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEY 424

```

```

Qy 425 GLAAGIHTSNINTALKVADRVNACTVWINTYNDFHHAVPFGGPNASGLGKREMSVVALONY 484
Db 420 GLSAAVHTSNINTALEVANALAGTVWVNSYNTLHHOMFEGGKESGIGRELGEDALANT 479
Qy 485 IAVKAVPAKL 494
Db 480 IOTKTVSIRL 489

RESULT 3
US-09-961-403-12
; Sequence 12, Application US/09961403
; Publication No. US20040077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLD
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,404
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-12

```

```

Query Match 47.8%; Score 1218.5; DB 9; Length 512;
Best Local Similarity 50.4%; Pred. No. 2,26-84;
Matches 240; Conservative 84; Mismatches 149; Indels 4; Gaps 4;

Qy 22 TGLFINNKVFPSKONKTFEVINISTEEELICHYEGREDVEFAVQAAURAFSNIS WNCI 80
Db 32 TRIFTNNEWHESKSGKKFATCNSTREQICEVEEGIKPVVDKAVDAVAVFQKSPWRRO 91
Qy 81 DPIDGKALYELAELEQIKIVIASIETLUNGKAISSRGIVREIVINYLESSAGAFAP 139
Db 92 DALSKGLLHQLADLVEROKALLALETMDLCKPELHAPFIDLERCTILKVFAGWAIKI 151
Qy 140 DGRMLITGTHESYTKKQPLAVQGIIPWNEPILMLWAKIAPALVTGNTVLKTAESTPL 199
Db 152 OGKTIPTDUNVVTFRHEPFGVSGALTPTWNTTLMVVKIADALASGIMVIRIADQID 211
Qy 200 SALYVSKYITPQAGIPPIVNIIVSGRGIIVFAIAELKKAQVNTILGIIYVNSGRVYAG 258
Db 212 TALYVSKYITPQAGIPPIVNIIVSGRGIIVFAIAELKKAQVNTILGIIYVNSGRVYAG 258
Qy 259 AGLKVVTELEGGKSNIVFAIAELKKAQVNTILGIIYVNSGRVYAGVYVEFSIYKFI 318
Db 272 SNLKVVTELEGGKSNIVFAIAELKKAQVNTILGIIYVNSGRVYAGVYVEFSIYKFI 341
Qy 319 EEFKAASESIKVGDPDEFDESTFQGAQTSOMQLNKLKLYVDIGKNEGA 378
Db 332 PFSVYAKKPPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEY 391
Qy 379 FIRKTPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEYGLIAGIHTSNINTA 438
Db 392 FIRKTPVGVKEDMPVIVKEEIFQPVVTVTKFKSADEVINMANSSEYGLIAGIHTSNINTA 438
Qy 439 IAVKAVPAKL 494
Db 452 IXLASALEFSVWINTYNAIYADAPFGGKESGIGRELGEDALANT 509

```

```

PFSHT 4
US-10-268-518-2
; Sequence 2, Application US/10268518

```



```

; APPLICANT: Lutziqer, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 2.2
; SEQ ID NO 20
; TYPE: PRT
; LENGTH: 538
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-20

```

```

Query Match      47.1% Score 1201; DB 9; Length 538;
Best Local Similarity 48.9% Pred. No. 5e-83;
Matches 233; Conservative 82; Mismatches 159; Indels 2; Gaps 2;

QY 22 TGLEINNKVFPSKNTKEFVINISTEEICHIEYEGREUVELAVUAAUKAFSNGSWNGID 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 TOLLINGNFVDSASGKTFTPTDPTGEVIAHVAESDAEDINRAVKAARTAFDEGWPVKMS 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 82 PIDRGKALYLAELIEDQKDVIASIEITLNGKAISSS-RGDVDLVINYLKSSAGFADKID 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 AYERSRVLLRFADLVEKHSEELASLETWNGKPYOOSLTAEIPMFARLFRYYAGWADKIH 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 141 GRMIDTGRTHFTYKROPFGVCGQIIPWNEPPLLMMAWKIAPALVTGNTVWLKTAESTPLS 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 GLTIPADGNTQVHTHEPIGVAGQIIPWNEPPLIMFAWKVGPALAGNTIVLKTATQTLT 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 201 ALYVSKYIPGAGIPPGVINTVSGFGKIVVEATINHPKIKKVAFTGSTATGRIHYOSAA-A 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 AFYAGKLFLEAGLPPGVNLIVSGFAGATAAALASHMVDKLAFTGSDTGKVIILGLAANS 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 260 GLKKVILELGKSPNIVFAELAKKAVQNLILGIIYNSGEVCCAGSRVYVEESIIDKFIIE 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 NLKPVTLLELGKSPFIIVFEDADIDKAVELAHFALFNQCCAGSRTPVHKYVDFEVE 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 320 EFKAASESIKVGPPDESTFGAGTSQMLNKLKYVDIGKNEGATLITGGERLGSKGYF 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 KSKARALKRVGDPFKKIEGQIQIDLKQFVKMYIKSGIESNATLEGQIQIDKGYF 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 380 IKPTFVDGDKMRIVKEEIEFGPVTVTKFKSADEVINMANDESEYGLAAGHTSNITNL 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 IQPTVESNVKDDMLIAQDEIFGPGVQSILKFSVDVEIKRANETKYGLAAGVETKNDLTAN 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 440 KVADRVNAGTVINTYNDHFHVAVPFGPGFNASGLGREGMSVDALONTYLOVKAVRAKID 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 RVSRALKAGTVVNCDFVDDAAIPFGGYKMSNGREKGIYSUNNYLQIKAVVVTALN 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
US-10-293-865-20
; Sequence 20, Application US/10/293865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziqer, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 217113

```

```

; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US 09/444,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 3.1
; SEQ ID NO 20
; TYPE: PRT
; LENGTH: 538
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-20

```

```

Query Match      47.1% Score 1201; DB 9; Length 538;
Best Local Similarity 48.9% Pred. No. 5e-83;
Matches 233; Conservative 82; Mismatches 159; Indels 2; Gaps 2;

QY 22 TGLEINNKVFPSKNTKEFVINISTEEICHIEYEGREUVELAVUAAUKAFSNGSWNGID 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 TOLLINGNFVDSASGKTFTPTDPTGEVIAHVAESDAEDINRAVKAARTAFDEGWPVKMS 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 82 PIDRGKALYLAELIEDQKDVIASIEITLNGKAISSS-RGDVDLVINYLKSSAGFADKID 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 AYERSRVLLRFADLVEKHSEELASLETWNGKPYOOSLTAEIPMFARLFRYYAGWADKIH 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 141 GRMIDTGRTHFTYKROPFGVCGQIIPWNEPPLLMMAWKIAPALVTGNTVWLKTAESTPLS 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 GLTIPADGNTQVHTHEPIGVAGQIIPWNEPPLIMFAWKVGPALAGNTIVLKTATQTLT 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 201 ALYVSKYIPGAGIPPGVINTVSGFGKIVVEATINHPKIKKVAFTGSTATGRIHYOSAA-A 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 AFYAGKLFLEAGLPPGVNLIVSGFAGATAAALASHMVDKLAFTGSDTGKVIILGLAANS 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 260 GLKKVILELGKSPNIVFAELAKKAVQNLILGIIYNSGEVCCAGSRVYVEESIIDKFIIE 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 NLKPVTLLELGKSPFIIVFEDADIDKAVELAHFALFNQCCAGSRTPVHKYVDFEVE 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 320 EFKAASESIKVGPPDESTFGAGTSQMLNKLKYVDIGKNEGATLITGGERLGSKGYF 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 KSKARALKRVGDPFKKIEGQIQIDLKQFVKMYIKSGIESNATLEGQIQIDKGYF 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 380 IKPTFVDGDKMRIVKEEIEFGPVTVTKFKSADEVINMANDESEYGLAAGHTSNITNL 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 IQPTVESNVKDDMLIAQDEIFGPGVQSILKFSVDVEIKRANETKYGLAAGVETKNDLTAN 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 440 KVADRVNAGTVINTYNDHFHVAVPFGPGFNASGLGREGMSVDALONTYLOVKAVRAKID 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 RVSRALKAGTVVNCDFVDDAAIPFGGYKMSNGREKGIYSUNNYLQIKAVVVTALN 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-09-344-882-24
; Sequence 24, Application US/09/444882
; Patent No. US20020162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziqer, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/444,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717

```


; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.2
; SEQ ID NO 24
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-24

Query Match 45.98; Score 1169.5; DB 9; Length 534;
Best Local Similarity 48.5%; Pred. No. 1.2e-80;
Matches 238; Conservative 77; Mismatches 167; Indels 9; Gaps 3;
QY 6 MTVPKIPNGLEYEQPTCLFINNKFPVSKONKTFEINPSTEEICHIVEGREDDVEAV 65
DB 45 ITPPVKV-----EHTQLLIGGRFVDVAVSGKFTPLDPPNGEVIAQVSGDAEDVNAV 97
QY 66 QAAAFASNSGWNITDIPDRCALYRLAELTETQKIVIASIETLDNCKAI-SSSRGVDVL 124
DB 98 AAARKADEGFWPKMTAYERSKILFPAULLIEKHNEIAALFTWNEKPYEQSAQIEVPM 157
QY 125 VINYLKSSAGPADKIDGPMIDTGTHTPSYTKPQDYGVCQGIIPWNEPFLMWAKIAPALV 184
DB 158 LARVFRYYAGWADKIHGTMTPGDGPHHVQTLHEPIGVAGQIIPWNEPFLMLSKLGPALA 217
QY 185 TGNIVLKTAEPTLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 244
DB 218 CGNIVVLKTAECTPLSALLVCKLLHEAGLPGGVNIVSGFGCAAGAAIAASHMLVDKVAFT 277
QY 245 GSTATGRHIYQSA-AGLKVYTLLELGGKSPNIVFADAEIKKAVQNIILGIIYNSGEVCCA 303
DB 278 GSTDVGKIIELASKSNLKVLTLEESHSFVCEADADVDQAVELAHFALFFNQGCCCA 337
QY 304 GSRVTVVERSIVDKLFEERKAASESIKYGDDEPDESTFOGAQTSOMQLNKILKYIDIGNEG 363
DB 338 GSRTFVHRVYDEVEREKAKALKNVGDGPFKSGIEGGQPOVDSEQFNKILKYIKHGEVAG 397
QY 364 ATLITGGRGSGYFIKPTVFGDVKEDMRIVKEEIPGPVTVTKFSADEVINMANDSE 423
DB 398 ATLGAQGRGSGYFIKPTVFGDVKEDMRIVKEEIPGPVTVTKFSADEVINMANDSE 457
QY 424 YGLAAGTHSNITALKVAHVNAIIVWNIYLFHFAVYFGGFGNAGSLGRKMSVDIALON 483
DB 458 YGLAAGVFTQNLDTAHLKALRVGIVWLNCFDVLASIPFGYKMSGIGREKGIYSLNN 517
QY 484 YLOVKAVRAKL 494
DB 518 YLOVKAVVTSL 528

RESULT 9
US-10-293-865-24
; Sequence 24, Application US/10293865
; Publication No. US20030106096A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinsan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/344,882
; PRIOR FILING DATE: 1998-06-25

; PRIOR APPLICATION NUMBER: US 60/060,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 24
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-24

Query Match 45.98; Score 1169.5; DB 9; Length 534;
Best Local Similarity 48.5%; Pred. No. 1.2e-80;
Matches 238; Conservative 77; Mismatches 167; Indels 9; Gaps 3;
QY 6 MTVPKIPNGLEYEQPTCLFINNKFPVSKONKTFEINPSTEEICHIVEGREDDVEAV 65
DB 45 ITPPVKV-----EHTQLLIGGRFVDVAVSGKFTPLDPPNGEVIAQVSGDAEDVNAV 97
QY 66 QAAAFASNSGWNITDIPDRCALYRLAELTETQKIVIASIETLDNCKAI-SSSRGVDVL 124
DB 98 AAARKADEGFWPKMTAYERSKILFPAULLIEKHNEIAALFTWNEKPYEQSAQIEVPM 157
QY 125 VINYLKSSAGPADKIDGPMIDTGTHTPSYTKPQDYGVCQGIIPWNEPFLMWAKIAPALV 184
DB 158 LARVFRYYAGWADKIHGTMTPGDGPHHVQTLHEPIGVAGQIIPWNEPFLMLSKLGPALA 217
QY 185 TGNIVLKTAEPTLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 244
DB 218 CGNIVVLKTAECTPLSALLVCKLLHEAGLPGGVNIVSGFGCAAGAAIAASHMLVDKVAFT 277
QY 245 GSTATGRHIYQSA-AGLKVYTLLELGGKSPNIVFADAEIKKAVQNIILGIIYNSGEVCCA 303
DB 278 GSTDVGKIIELASKSNLKVLTLEESHSFVCEADADVDQAVELAHFALFFNQGCCCA 337
QY 304 GSRVTVVERSIVDKLFEERKAASESIKYGDDEPDESTFOGAQTSOMQLNKILKYIDIGNEG 363
DB 338 GSRTFVHRVYDEVEREKAKALKNVGDGPFKSGIEGGQPOVDSEQFNKILKYIKHGEVAG 397
QY 364 ATLITGGRGSGYFIKPTVFGDVKEDMRIVKEEIPGPVTVTKFSADEVINMANDSE 423
DB 398 ATLGAQGRGSGYFIKPTVFGDVKEDMRIVKEEIPGPVTVTKFSADEVINMANDSE 457
QY 424 YGLAAGTHSNITALKVAHVNAIIVWNIYLFHFAVYFGGFGNAGSLGRKMSVDIALON 483
DB 458 YGLAAGVFTQNLDTAHLKALRVGIVWLNCFDVLASIPFGYKMSGIGREKGIYSLNN 517
QY 484 YLOVKAVRAKL 494
DB 518 YLOVKAVVTSL 528

RESULT 10
US 09 344 882-23
; Sequence 22, Application US/09344882
; Patent No. US20030137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinsan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/060,717

; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Patent In Ver. 2.2
 ; SEQ ID NO 22
 ; LENGTH: 501
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis Thaliana
 US-09-344-882-22

Query Match 45.0%; Score 1146.5; DH 9; Length 501;
 Best Local Similarity 47.1%; Pred. No. 6.2e-79;
 Matches 228; Conservative 95; Mismatches 154; Indels 7; Gaps 5.

```

QY 10 IKLPNGLEYEQPTGLFINNKVFPSKUNKTFEVINESIEIECHIVIEGREDDEVERAVQAAD 69
DB 12 VKLPE-IKF---TKLFINGQFIDAAAGKTFETIDPRNGEVIATIAEGKEDVDIAVNAAR 67
QY 70 RAFNSGSMGIDPIDRGKALRYLAELIEQDKOVIASIETLDNGKAISSR-GDVLVINY 128
DB 68 YAFDHGWPWMTGFERAKLINKEFADLIEENIEELAKLDADVGGKLFQLGKYADIPATAGH 127
QY 129 LKSSAGFADKIDGRMID-TGRTHEFSYTKRQPLGVCGGIIIPWNPFLLMWAKKIPALVTGN 187
DB 128 FRYNAGAADKIHGKTLKMTQSLFGVTLKEPIGVVGNIIIPWNPFSIMFATKVAPAMAAGC 187
QY 188 TVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFKIVVEATINHPKIKKVAFTGST 247
DB 188 TMVVKPAEQTSLSALFYAHLSEKAGIPDGVNLIVTGFSTAGAAIASHMDVDVKVFTGST 247
QY 248 ATGHHIVQSAAA-GLKKVILELQSKSPNIVFAIAELKKAQVNIILGIIYNSGEVCCASR 306
DB 248 LVGRKIMCAAAA-NLKKVILELQSKSPNIVFAIAELKKAQVNIILGIIYNSGEVCCASR 307
QY 307 VYVEESIYDKFIEFEKAASESIRKVGDPFDESTFOGAQTSOMOLNKILKYVDITGKNEGATL 366
DB 308 VFVQEGYDKVWEKLVKAKUMTVGDPFDSIAKQGVVUKKQFEKILSYIEGKNEGAIL 367
QY 367 ITGGERIGSGYFIKPTVFGDVEDMKRIYKEEIFGVPVTVTKFSADADEVINMANSEYGL 426
DB 368 LTGCKAIGDKGYFIQPTIFADVTEDMKIYQDEIFGVPVMSLMKFEKTVESGKCANNTKYL 427
QY 427 AAGLHTSNINTALKVADRVNAGTVWINTYNDHFHAPVPGSGNAGSLGKMSVDALQNYLO 486
DB 428 AAGILSQDIDLINTVSRSKAGIIWNVYEGFELDCTPGGYKMSGNCRSGMDALDNYLO 487
QY 487 VKAV 490
DB 488 TKSV 491
  
```

RESULT 11

US-10-293-865-22
 ; Sequence 22, Application US/10293865
 ; Publication No. US20030106090A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nikolau, Basil J
 ; APPLICANT: Wurtele, Eve S
 ; APPLICANT: Oliver, David J
 ; APPLICANT: Behal, Robert
 ; APPLICANT: Schenable, Patrick S
 ; APPLICANT: Ke, Jinshan
 ; APPLICANT: Johnson, Terry L
 ; APPLICANT: Allred, Carolyn C
 ; APPLICANT: Fatland, Beth
 ; APPLICANT: Lutziger, Isabelle
 ; APPLICANT: Wen, Tsui-Jung
 ; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 ; FILE REFERENCE: 217113
 ; CURRENT APPLICATION NUMBER: US/10/293,865
 ; PRIOR FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/444,882
 ; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 642/000,713
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Patent In Ver. 3.1
 ; SEQ ID NO 22
 ; LENGTH: 501
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis Thaliana
 US-10-293-865-22

Query Match 45.0%; Score 1146.5; DH 9; Length 501;
 Best Local Similarity 47.1%; Pred. No. 6.2e-79;
 Matches 228; Conservative 95; Mismatches 154; Indels 7; Gaps 5.

```

QY 10 IKLPNGLEYEQPTGLFINNKVFPSKUNKTFEVINESIEIECHIVIEGREDDEVERAVQAAD 69
DB 12 VKLPE-IKF---TKLFINGQFIDAAAGKTFETIDPRNGEVIATIAEGKEDVDIAVNAAR 67
QY 70 RAFNSGSMGIDPIDRGKALRYLAELIEQDKOVIASIETLDNGKAISSR-GDVLVINY 128
DB 68 YAFDHGWPWMTGFERAKLINKEFADLIEENIEELAKLDADVGGKLFQLGKYADIPATAGH 127
QY 129 LKSSAGFADKIDGRMID-TGRTHEFSYTKRQPLGVCGGIIIPWNPFLLMWAKKIPALVTGN 187
DB 128 FRYNAGAADKIHGKTLKMTQSLFGVTLKEPIGVVGNIIIPWNPFSIMFATKVAPAMAAGC 187
QY 188 TVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFKIVVEATINHPKIKKVAFTGST 247
DB 188 TMVVKPAEQTSLSALFYAHLSEKAGIPDGVNLIVTGFSTAGAAIASHMDVDVKVFTGST 247
QY 248 ATGHHIVQSAAA-GLKKVILELQSKSPNIVFAIAELKKAQVNIILGIIYNSGEVCCASR 306
DB 248 LVGRKIMCAAAA-NLKKVILELQSKSPNIVFAIAELKKAQVNIILGIIYNSGEVCCASR 307
QY 307 VYVEESIYDKFIEFEKAASESIRKVGDPFDESTFOGAQTSOMOLNKILKYVDITGKNEGATL 366
DB 308 VFVQEGYDKVWEKLVKAKUMTVGDPFDSIAKQGVVUKKQFEKILSYIEGKNEGAIL 367
QY 367 ITGGERIGSGYFIKPTVFGDVEDMKRIYKEEIFGVPVTVTKFSADADEVINMANSEYGL 426
DB 368 LTGCKAIGDKGYFIQPTIFADVTEDMKIYQDEIFGVPVMSLMKFEKTVESGKCANNTKYL 427
QY 427 AAGLHTSNINTALKVADRVNAGTVWINTYNDHFHAPVPGSGNAGSLGKMSVDALQNYLO 486
DB 428 AAGILSQDIDLINTVSRSKAGIIWNVYEGFELDCTPGGYKMSGNCRSGMDALDNYLO 487
QY 487 VKAV 490
DB 488 TKSV 491
  
```

RESULT 12

US-10-268-518-4
 ; Sequence 4, Application US/10268518
 ; Publication No. US20040100044A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John Joseph
 ; TITLE OF INVENTION: 9136, A HUMAN ALBANYDE DEHYDRONASE
 ; FILE REFERENCE: M101-244PIRM
 ; CURRENT APPLICATION NUMBER: US/10/268,518
 ; PRIOR FILING DATE: 2002-10-10
 ; PRIOR APPLICATION NUMBER: 60/429,899
 ; PRIOR FILING DATE: 2001-10-16
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Consensus sequence
 US-10-268-518-4

Query Match 41.88; Score 1067; DR 9; Length 492;
 Best Local Similarity 46.98; Pred No 6, file 74;
 Matches 231; Conservative 98; Mismatches 136; Indels 28; Gaps 15;

```

QY 29 KFPVSKNKTPEVNPSTEEELIC-HIYEGHEDUVEEAVGAADKAPFNGSNWIDPT-DPG 86
DB 1 EWDUSASCKTPEVNPANKGEVIGRVPEATAEDUADAAVKAKEAFKSGPWAKVPASEKA 60
QY 87 KALYELAELEIQQDKVIASTIELONGKAISSSGERVEL--VINYLKSSAPLAKING--R 142
DB 61 RTIPKIALDIEBERDELAALLETIDLKPLAEAKGTEVSGPAIDEIHYAGWAPKLMGPRP 120
QY 143 MIDT-----GRTHFSYTRQPIGVGGQIIPWNEFLMKWAKIAPALVTGNTVLKTAESTP 198
DB 121 VPSLAIDGDEELNYTPPEPISVWVVISPNWPFILITAIWKLAIAIAGNTIVIKPSQGP 180
QY 199 LSALYVSKYIPJAG---IPGVINIVSGFGKIIVVEALINHPKIKKVAFTJSTATGPHIVQ 255
DB 181 LTALLAELEIEAGANNLPKGVNVVPGFAEVGQALLSHPDIDKISFTGSTEVGKLINE 240
QY 256 SAAA-GLKKVILELGGKSPNIVFADAEIKKAVGNILIGIYNSGFWCCAGSPVYVEESY 314
DB 241 AAAAKNKKVILELGGKSPNIVFADAEIKKAVGNILIGIYNSGFWCCAGSPVYVEESY 300
QY 315 DKFIEEFKAASESIK-VSDPPEDEST-FQSAQISQMLNKLIL-KYVDIGKNEGATLITGGE 371
DB 301 DEFPVKLEKPKVKIKIIGDPIISLFINIYGPILSGQGFQVWSYIFDGRKEAKVLGSGE 360
QY 373 LSSK-----GYFTKPTVFGVEKEMPVKKEEFISPVVIVTKRKALEVINMANDESYHIA 427
DB 361 DESKPYIGGGYVQGFPIITVTPUMKIMKEEFISPVVILELKKKQDLEALIELANDTEYGLA 420
QY 428 AGIHTSNI-NTALKVADRVNAGTIVWINTYNGFHHA---VDFGSE-NASGLGSEMSVLA-AL 481
DB 421 AAYVETKIDILAKAFKAVAKALEAIVWVNVIV-CVHAAPFQIPFGGVKSGSSIGSEHSGKYG 479
QY 482 QNYLQVKAVRAKL 494
DB 480 LEYTEIKTIVIRL 493

```

RESULT 13

```

US-09-823-901-9
; Sequence 21, Application US/10175696
; Patent No. US2001009265A1
; GENERAL INFORMATION:
; APPLICANT: Glücksmann, Maria Alexandra
; APPLICANT: Glücksmann, Rachel
; TITLE OF INVENTION: NOVEL HUMAN ENZYME FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-193001
; CURRENT APPLICATION NUMBER: US/10175,696
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/067,668
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/823,901
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: PCT/US01/10720
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/193,920
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/862,658
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16360
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/205,675
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/882,837
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19319

```

```

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 06/211,727
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 493
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: cons-nus sequence
us-10-175-696-21

```

Query Match 41.48; Score 1056.5; DR 9; Length 493;
 Best Local Similarity 46.88; Pred No 4, file 72;
 Matches 231; Conservative 98; Mismatches 136; Indels 29; Gaps 16;

```

QY 29 KFPVSKNKTPEVNPSTEEELIC-HIYEGHEDUVEEAVGAADKAPFNGSNWIDPT-DPG 86
DB 1 EWDUSASCKTPEVNPANKGEVIGRVPEATAEDUADAAVKAKEAFKSGPWAKVPASEKA 60
QY 87 KALYELAELEIQQDKVIASTIELONGKAISSSGERVEL--VINYLKSSAPLAKING--R 142
DB 61 RTIPKIALDIEBERDELAALLETIDLKPLAEAKGTEVSGPAIDEIHYAGWAPKLMGPRP 120
QY 143 MIDT-----GRTHFSYTRQPIGVGGQIIPWNEFLMKWAKIAPALVTGNTVLKTAESTP 198
DB 121 VPSLAIDGDEELNYTPPEPISVWVVISPNWPFILITAIWKLAIAIAGNTIVIKPSQGP 180
QY 199 LSALYVSKYIPJAG---IPGVINIVSGFGKIIVVEALINHPKIKKVAFTJSTATGPHIVQ 255
DB 181 LTALLAELEIEAGANNLPKGVNVVPGFAEVGQALLSHPDIDKISFTGSTEVGKLINE 240
QY 256 SAAA-GLKKVILELGGKSPNIVFADAEIKKAVGNILIGIYNSGFWCCAGSPVYVEESY 314
DB 241 AAAAKNKKVILELGGKSPNIVFADAEIKKAVGNILIGIYNSGFWCCAGSPVYVEESY 300
QY 315 DKFIEEFKAASESIK-VSDPPEDEST-FQSAQISQMLNKLIL-KYVDIGKNEGATLITGGE 371
DB 301 DEFPVKLEKPKVKIKIIGDPIISLFINIYGPILSGQGFQVWSYIFDGRKEAKVLGSGE 360
QY 372 PLGSK-----GYFTKPTVFGVEKEMPVKKEEFISPVVIVTKRKALEVINMANDESYL 426
DB 361 DESKPYIGGGYVQGFPIITVTPUMKIMKEEFISPVVILELKKKQDLEALIELANDTEYGL 420
QY 427 AAGIHTSNI-NTALKVADRVNAGTIVWINTYNGFHHA---VDFGSE-NASGLGSEMSVLA-AL 480
DB 421 AAYVETKIDILAKAFKAVAKALEAIVWVNVIV-CVHAAPFQIPFGGVKSGSSIGSEHSGKYG 479
QY 481 QNYLQVKAVRAKL 494
DB 480 LEYTEIKTIVIRL 493

```

RESULT 14

```

US-09-823-901-9
; Sequence 9, Application US/09823901
; Patent No. US2002001807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 21509 AND 33770, NOVEL HUMAN
; TITLE OF INVENTION: DEHYDROGENASES AND USES THEREOF
; FILE REFERENCE: 10448-036001
; CURRENT APPLICATION NUMBER: US/09/823,901
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/862,658
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/193,920
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 493
; TYPE: PPT
; OTHER INFORMATION: cons-nus sequence

```

```

; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-823-901-9

Query Match      41.4%; Score 1056.5; DB 10; Length 493;
Best local Similarity 46.8%; Pred No. 4 10-72;
Matches 231; Conservative 98; Mismatches 136; Indels 29; Gaps 16;

QY 29 KFYPSKONKTEVINSTEELIC-HIYEGRELDVEAVQAALAHAFNSNSWNGIDPI-DKG 86
Db 1 EWDSASGKTEFFVNPANKEVIGKVPATEADVAAVKAAKAFKSGPWWAKVPASRPA 60

QY 87 KALYRLAELIEQKQDVIASTETLDNGKATISSRGDVL--VINYLKSSAGFADKIG--R 142
Db 61 RILRLADLEEDPEDELALETLDLGGKPLAPAKGTEVCRALDELRYVAGHAPKLMGERP 120

QY 143 MIDT----GRTHSYTKRPLGVCGGQIPWNEPPLLMWAKIAPALVTGNTVYLKTAESTP 198
Db 121 VIPSLATDGEELNTRREPLGVGVISPWNEPPLLLALWKLAPALAAAGNTVYLKPEQTP 180

QY 199 LSALYYSKYIPOAG---IPPGVINIVSGGKTVVFAITNHPKIKKVAFTGSTATCPHIYQ 255
Db 181 LTFALLAELIEAGANLKPQVNVVPGFAGVGGALLSHPDIDKISFTGSTEVKLTME 240

QY 256 SAAA-GLKKVTLLELGGKSNIVFADAEALKKAVONIILGIYINSGEVCCAGSRVYVERSIY 314
Db 241 AAAAKNLKVTLELGGKSPVIVEDDADLLKAVERTVFAFGNAQVCTAPSKLLVHESIY 300

QY 315 DKFIEFKAESIK-VQPPFEST-FQCAQTSQMLNKII-KYVDIGKNEGATLITGGE 371
Db 301 DEVEKLEKRVKKLIGLPGDSQINIVGPIISEQGFQKRVLSYEDGKPKGAKVLDGGE 360

QY 372 FJASK--GYFKPTVGLGKQEMFIVKKEIF-PPVIVIVIKESAGEVINMAN-SEYGL 426
Db 361 RDSKEYLGYYVQPTIFTDTVPDKIMKEIFGVPVPIIKPKDLDEAIELANDTEYGL 420

QY 427 AAGIHTSNI-NITALKVADVAGNAGTVWINTYNDFHHA---VPEGGF-NASGLGKREMSVD-A 480
Db 421 AAVFTKDLARAFRAVKALEAGIVWVNDV-CVHAAPQLPFGGVKQSSGIGREHGKYG 479

QY 481 LONYLOVKAVRAKL 494
Db 480 LEYTEIKTVITKL 493

RESULT 15
US-09-815-242-10550
; Sequence 10550, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

```

; PRIOR APPLICATION NUMBER: 60/257,941
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10550
; LENGTH: 496
; TYPE: PPT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10550

Query Match      40.2%; Score 1025; DB 10; Length 496;
Best local Similarity 43.6%; Pred No. 10-69;
Matches 206; Conservative 87; Mismatches 171; Indels 8; Gaps 4;

QY 24 LFINNKVPSKONKTEVINSTEELICHIYEGRELDVEAVQAADRAFNSNSWNGIDPI 83
Db 16 LYINGFWPTGSGNKMFIASVNSCKELAFEVVATINAVIHRAVFAAQAFQ 1KKQVIVV 74

QY 84 DPKKALYRLAELIEQKQDVIASTETLDNGKATISSRGDVL--VINYLKSSAGFADKIG--R 142
Db 74 TRSNLLKRTADLIEENQEHILAMVETLDNGKPIRETQSIDVDPASAIHIFYPASVIRGEES 144

QY 143 MIDTGRTHSYTKRPLGVCGGQIPWNEPPLLMWAKIAPALVTGNTVYLKTAESTP 202
Db 144 VKFEFKDILSVVKEPDAVVGQIPWNEPPLLMWAKIAPALAAVNTVYIHSSSTLSLL 194

QY 203 YYSKYIPOAG---IPPGVINIVSGGKTVVFAITNHPKIKKVAFTGSTATCPHIYQ 252
Db 194 ELFKTFDAV-LPKVWVNLITGKSSNSNYMLAHPTGFKLAFPTSEVTVYVAKAAADRL 252

QY 253 KVLLELGGKSNIVFADAEALKKAVONIILGIYINSGEVCCAGSRVYVERSIY 314
Db 253 PATLELGGKSNIIPEFADANERALEGCVOLGILFNGQCVVACGIVTVQSGIYVQVIALK 312

QY 323 AASESIKVGDPDESTFQCAQTSQMLNKII-KYVDIGKNEGATLITGGERIAG--SKGY 378
Db 313 EKFEQVNVGFPWEKQVEMCAQINEHQLEILLKYVEIGVKKECATLITGQRETNELGKA 372

QY 379 FKPTVFGDVKEDMKRIVKEEIFGPPVTVTKPSADEVINMANDEYGLAAGIHTSNI 448
Db 373 FLAPTLANGTNTMCVAQETPGVATVTKPETEVEVIRLANDSEYGLGAVFSQDINVA 442

QY 439 LKVADVAGNAGTVWINTYNDFHHAIVPFGFNAGSLGKREMSVDIALQNYQVKA 490
Db 434 LKVAAGVPLGPPWVWNTYNGILFAGALGKSSGLGKPKFTHKSMIYAYTGMPKI 484

```

Search completed: June 24, 2003, 10:47:03
Job time : 24.9275 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:11:00 : Search time 174.237 seconds
(without alignments)
1780 862 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSRLPMIVPIKPLNGLEYEQ... MSVFALNYLYKAVPAKILN 495

Scoring table: PROSNM62
Gapop 10 0 , Gapext 0.5

Searched: 4569144 seqs, 6447 x110 residues 4569144
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database		Pending_Patents_AA_Main	
1:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*	1:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*	2:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*	3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	4:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	5:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	6:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	7:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	8:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	9:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	10:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	11:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	12:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	13:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	14:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	15:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	16:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	17:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	18:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	19:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	20:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	21:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	22:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	23:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*	24:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*	25:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*	26:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*	27:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2538	99.5	519	21	US-09-791-537-47094, A
2	2538	99.5	519	27	US-60-360-039-22438, A
3	1807	70.9	518	21	US-09-791-537-38066, A
4	1728.5	67.8	520	27	US-09-791-537-19395, A
5	1728.5	67.8	520	27	US-60-360-039-17499, A
6	1696	64.5	519	21	US-09-791-537-70112, A

7	1621.5	63.6	519	25	US-10-179-131-8239, A
8	1455	57.1	447	31	US-09-791-537-24771, A
9	1452	56.9	509	21	US-09-791-537-61775, A
10	1452	56.9	509	27	US-60-360-039-22495, A
11	1443.5	56.6	501	21	US-09-791-537-70619, A
12	1442.5	56.2	448	27	US-60-360-039-4846, A
13	1412.5	55.4	497	10	US-08-612-853-11, A
14	1412.5	55.4	497	21	US-09-791-537-73644, A
15	1399	54.9	496	10	US-08-612-853-2, A
16	1399	54.9	496	21	US-09-791-537-9158, A
17	1399	54.9	496	22	US-09-847-208-59, A
18	1399	54.9	496	22	US-09-847-208B-59, A
19	1394	54.7	496	10	US-08-612-853-3, A
20	1394	54.7	497	10	US-08-612-853-10, A
21	1362.5	53.4	444	16	US-09-791-537-26770, A
22	1362.5	53.4	444	16	US-09-791-537-26770, A
23	1362.5	53.4	495	22	US-09-847-208-11, A
24	1362.5	53.4	495	22	US-09-847-208B-11, A
25	1362.5	53.4	496	10	US-08-612-853-9, A
26	1362.5	53.4	503	21	US-09-791-537-115973, A
27	1325.5	52.3	427	11	US-09-791-537-21613, A
28	1297	50.9	382	16	US-09-348-746-17019, A
29	1297	50.9	382	27	US-60-360-039-17019, A
30	1275.5	50.0	519	21	US-09-791-537-16904, A
31	1274.5	50.0	519	21	US-09-791-537-74080, A
32	1265	49.5	539	25	US-10-179-141-5478, A
33	1261.5	49.5	500	21	US-09-791-537-64443, A
34	1256.5	49.4	517	27	US-60-360-039-1227, A
35	1256.5	49.3	517	27	US-60-412-418-1227, A
36	1254.5	49.2	532	27	US-60-360-039-12695, A
37	1253.5	49.2	517	19	US-09-538-092-851, A
38	1253.5	49.2	517	21	US-09-791-537-74079, A
39	1253.5	49.2	517	27	US-60-389-987-1242, A
40	1253.5	49.2	517	27	US-60-412-418-1242, A
41	1252.5	49.1	444	21	US-09-791-537-4937, A
42	1252.5	49.1	495	27	US-60-360-039-23162, A
43	1252.5	49.1	517	21	US-09-791-537-60422, A
44	1251.5	49.1	494	21	US-09-791-537-66667, A
45	1250.5	49.0	517	21	US-09-791-537-71773, A

ALIGNMENTS

RESULT 1
US-09-791-537-47094 Application US/09791537
Sequence 47094
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
INVENTOR: Dwyer, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER US/09791537
CURRENT FILING DATE: 2001 02 22
NUMBER OF SEQ IDS: 154055
SOFTWARE: Patent in version 3.0
SEQ ID NO 47094
LENGTH: 519
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-791-537-47094

Query Match	99.5%	Score	2538	DR 21	Length	519			
Best Local Similarity	99.88%	Pred. No.	5e+230						
Matches	493	Conservative	0	Mismatches	1	Indels	0	Gaps	0
Q1	2	SHLEMTVPIKPLNGLEYEQTGLFINKFVPSKNTFEVINSTEETFECHVSEEDVV	61						
Q2	25	SHLEMTVPIKPLNGLEYEQTGLFINKFVPSKNTFEVINSTEETFECHVSEEDVV	84						
Q3	42	SHLEMTVPIKPLNGLEYEQTGLFINKFVPSKNTFEVINSTEETFECHVSEEDVV	121						

```

Db      85  EEAVQADRAAFSNQWNGIDPDRKALYPLAFI IQQKDVFIASITLNGKALSSSGD 144
QY      122  VDLVINYLKSSAGFAIDKIDGRMIDTGRTHFSYTKQPLGVCGQIIPWNPPLLMWAKIAP 181
Db      145  VDLVINYLKSSAGFAIDKIDGRMIDTGRTHFSYTKQPLGVCGQIIPWNPPLLMWAKIAP 204
QY      182  ALVTGNTVWLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 241
Db      205  ALVTGNTVWLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 264
QY      242  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 301
Db      265  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 324
QY      302  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 361
Db      325  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 384
QY      362  EGATLTITGGRLGSKGYFIKPTVFGDVKEDMRIVKEEIEGPPVTVTKKSADEVINMAMD 421
Db      385  EGATLTITGGRLGSKGYFIKPTVFGDVKEDMRIVKEEIEGPPVTVTKKSADEVINMAMD 444
QY      422  SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 481
Db      445  SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 504
QY      482  QNYLOVKAVRAKLD 495
Db      505  QNYLOVKAVRAKLD 518

```

RESULT 2

US-60-360-039-224 48

Sequence 22438, Application US/60360039

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Chen, Xianfeng

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)A

CURRENT APPLICATION NUMBER: US/60/360,039

CURRENT FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 22438

LENGTH: 519

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-60-360-039-224 48

```

Query Match      99.8%, Score 2538, FR 27, Length 519.
Best Local Similarity 99.8%, Pred. No. 5e-2307,
Matches 493, Conservative 3, Mismatches 1, Indels 6, Gaps 0:

```

```

QY      2  SHLPMTVFIKLPNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEIECHYIEGREDV 61
Db      25  SHLPMTVFIKLPNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEIECHYIEGREDV 84
QY      62  EEAVQADRAAFSNQWNGIDPDRKALYPLAFI IQQKDVFIASITLNGKALSSSGD 121
Db      85  EEAVQADRAAFSNQWNGIDPDRKALYPLAFI IQQKDVFIASITLNGKALSSSGD 144
QY      122  VDLVINYLKSSAGFAIDKIDGRMIDTGRTHFSYTKQPLGVCGQIIPWNPPLLMWAKIAP 181
Db      145  VDLVINYLKSSAGFAIDKIDGRMIDTGRTHFSYTKQPLGVCGQIIPWNPPLLMWAKIAP 204
QY      182  ALVTGNTVWLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 241
Db      205  ALVTGNTVWLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 264

```

```

QY      242  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 301
Db      265  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 324
QY      302  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 361
Db      325  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 384
QY      362  EGATLTITGGRLGSKGYFIKPTVFGDVKEDMRIVKEEIEGPPVTVTKKSADEVINMAMD 421
Db      385  EGATLTITGGRLGSKGYFIKPTVFGDVKEDMRIVKEEIEGPPVTVTKKSADEVINMAMD 444
QY      422  SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 481
Db      445  SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 504
QY      482  QNYLOVKAVRAKLD 495
Db      505  QNYLOVKAVRAKLD 518

```

RESULT 3

US-09-791-537-48066

Sequence 38066, Application US/09791547

GENERAL INFORMATION:

APPLICANT: Bionomix, Inc.

APPLICANT: Debes, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

TITLE OF INVENTION: METHODS OF USE THEREOF

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09/791,547

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: Patent In version 4.0

SEQ ID NO 38066

LENGTH: 518

TYPE: PRT

ORGANISM: Pichia ananasta

US-09-791-547-48066

Query Match 70.9%, Score 1807, FR 21, Length 518.

Best Local Similarity 68.2%, Pred. No. 5.9e-161,

Matches 337, Conservative 72, Mismatches 81, Indels 2, Gaps 2:

```

QY      2  SHLPMTVFIKLPNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEIECHYIEGREDV 61
Db      22  SSYPISITVTLNGSKSYEQTGLFINNKEVPSKQNKTEFVNPSTEEIECHYIEGREDV 81
QY      62  EEAVQADRAAFSNQWNGIDPDRKALYPLAFI IQQKDVFIASITLNGKALSSSGD 121
Db      82  DWAVDAAEKAF KGSWSTALPAIRGALWNLAELEAIKDTLAALESLENGKALQAQD 140
QY      122  VDLVINYLKSSAGFAIDKIDGRMIDTGRTHFSYTKQPLGVCGQIIPWNPPLLMWAKIAP 181
Db      141  VALVINTFSCAWALKLGGYTDGDDHDTAREPQWGGQITPWFPELMSWAKIAP 200
QY      182  ALVTGNTVWLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 241
Db      201  ALAA:NTVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIIVVEAITNHPKIKKY 260
QY      242  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 300
Db      261  AFTGSTATGRHLYQSAAGLKKVTLLEGGKSPNIVFADAEELKKAVQNIILGLIYNSGEVC 320
QY      301  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 360
Db      321  CAGSRVYVESIYDKFIEEPKAASESIKVGDPDESTFOGAOTSOMQLNKILKYVDIGKN 380
QY      361  NEVAILIDGEPDSSNGYIEKIVFVVEEDMPVFEETFPVVIIVFEFSADDEVINMAN 420
Db      481  QNYLOVKAVRAKLD 495

```

QY 421 DSEYGLAAGIHTSNINTALKVADRYNAGTVMINTYNDPHHAYPFGGPNASGLGEMSVDA 480
 Db 441 DSEYGLAAGVHTKSLDTATYVSNKLEAGSVMTYNDPHOMVPPGGYKQSGIGREMGQA 500
 QY 481 LQNYLQVKAVRAKL 494
 Db 501 LQNYTQWKAIRGL 514

RESULT 4

US-09-791-537-70612

Sequence 105395, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Blomox, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: Patent in version 3.0

SEQ ID NO 105395

LENGTH: 520

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-791-537-70612

Query Match 67.8%, Score 1728.5, DB 21, Length 520;
 Best Local Similarity 64.8%, Pred. No 16e-153;
 Matches 320; Conservative 81; Mismatches 92; Indels 1, Gaps 1,

QY 2 SHLPMTVPKIPNGLEYEQPTGLFINNKFPVSKONKTFEIVNIPSTEEICHIEGREDVV 61
 Db 24 SQAPLRVPITLPGNGFTYEQPTGLFNGEFVASKOKKTFDVINFSNEEKITIVYKAMEDDV 83

QY 62 EAVVAADRAFNSGWSNGIDPDRGKALYRLAELIQKIVIASTETLNGKALSSRGD 121
 Db 84 DEAVAAAKAFET-KWSIVEPEVRKALFNLAIVKHOETLAAIESMDNGKSLFCARGD 142

QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPNFPLLMMAWKIAP 181
 Db 143 VALVSKYLRSCGGWADKIYGNVDTGKNHFTYSIKEPLGVCGQIIPNFPLLMMSWKIGP 202

QY 182 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINISYGGKIVVEAITNHPKTKV 241
 Db 203 ALATGNTVVLKPAETITPLSALFASQLOEAGIPACVNVNIPGSRVVGRELSAHPDVYKI 262

QY 242 AFTGSTATGRHLYQSAAGLKKVTLELGGKSPNIVFAELKAVQNIILGIYNSGEVC 301
 Db 263 AFTGSTATGRHLMKVAADTVKVTLELGGKSPNIVFAEADLDKAVKNIAFGIYNSGEVC 322

QY 302 CAGSRVYVEESTYDKFIEEFRAASHSIKVGDPPFDETFQGAQTSQMLNKLKLYVDIGKN 361
 Db 323 CAGSRVYIQDTVYEPVIAKLYTESLKVKV-LPPGFEVFEAGATSIKGLIKLIDYVDVAKS 382

QY 362 EGATLITGGERLGSKGYSYFKPTVTRGVKEMFIVKFEIFPVTVTFKFSADDEVINMAD 421
 Db 383 EGARLVTSGARHSGKGYEVKPTVFADVKGMRIYKEEVFGPIVTSKFSSTVDEVIANMAD 442

QY 422 SEYGLAAGIHTSNINTALKVADRYNAGTVMINTYNDPHHAYPFGGPNASGLGEMSVDA 481
 Db 443 SQYGLAAGIHTNDINKAVDSKRVKAGTIVTNNFHNQVPPFGGFGSGIGREMGQAAL 502

QY 482 QNYLQVKAVRAKL 495
 Db 503 SNTYQTKSVRIAID 516

RESULT 5

US-60-360-039-1749

Sequence 1749, Application US/60360039

GENERAL INFORMATION:

APPLICANT: Blomox, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: Patent in version 3.0

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Chen, Xianfeng
 APPLICANT: Goldman, Barry S.
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)A
 CURRENT APPLICATION NUMBER: US/60/360,039
 CURRENT FILING DATE: 2002 02 21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 1749
 LENGTH: 520
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-60-360-039-1749

Query Match 67.8%, Score 1728.5, DB 27, Length 520;
 Best Local Similarity 64.8%, Pred. No 1.6e-153;
 Matches 320; Conservative 81; Mismatches 92; Indels 1; Gaps 1;

QY 2 SHLPMTVPKIPNGLEYEQPTGLFINNKFPVSKONKTFEIVNIPSTEEICHIEGREDVV 61
 Db 24 SQAPLRVPITLPGNGFTYEQPTGLFNGEFVASKOKKTFDVINFSNEEKITIVYKAMEDDV 83

QY 62 EAVVAADRAFNSGWSNGIDPDRGKALYRLAELIQKIVIASTETLNGKALSSRGD 121
 Db 84 DEAVAAAKAFET-KWSIVEPEVRKALFNLAIVKHOETLAAIESMDNGKSLFCARGD 142

QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPNFPLLMMAWKIAP 181
 Db 143 VALVSKYLRSCGGWADKIYGNVDTGKNHFTYSIKEPLGVCGQIIPNFPLLMMSWKIGP 202

QY 182 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINISYGGKIVVEAITNHPKTKV 241
 Db 203 ALATGNTVVLKPAETITPLSALFASQLOEAGIPACVNVNIPGSRVVGRELSAHPDVYKI 262

QY 242 AFTGSTATGRHLYQSAAGLKKVTLELGGKSPNIVFAELKAVQNIILGIYNSGEVC 301
 Db 263 AFTGSTATGRHLMKVAADTVKVTLELGGKSPNIVFAEADLDKAVKNIAFGIYNSGEVC 322

QY 302 CAGSRVYVEESTYDKFIEEFRAASHSIKVGDPPFDETFQGAQTSQMLNKLKLYVDIGKN 361
 Db 323 CAGSRVYIQDTVYEPVIAKLYTESLKVKV-LPPGFEVFEAGATSIKGLIKLIDYVDVAKS 382

QY 362 EGATLITGGERLGSKGYSYFKPTVTRGVKEMFIVKFEIFPVTVTFKFSADDEVINMAD 421
 Db 383 EGARLVTSGARHSGKGYEVKPTVFADVKGMRIYKEEVFGPIVTSKFSSTVDEVIANMAD 442

QY 422 SEYGLAAGIHTSNINTALKVADRYNAGTVMINTYNDPHHAYPFGGPNASGLGEMSVDA 481
 Db 443 SQYGLAAGIHTNDINKAVDSKRVKAGTIVTNNFHNQVPPFGGFGSGIGREMGQAAL 502

QY 482 QNYLQVKAVRAKL 495
 Db 503 SNTYQTKSVRIAID 516

RESULT 6

US-09-791-537-70612

Sequence 70612, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Blomox, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: Patent in version 3.0

Db 485 EVVHAYTEVAVRIKL 500

RESULT 11
US-09-791-537-70619
Sequence 70619, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 70619
LENGTH: 501
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-791-537-70619

Query Match 56.6%; Score 1443.5; DB 21; Length 501;
Best Local Similarity 54.7%; Pred. No. 1,3e-126;
Matches 272; Conservative 92; Mismatches 128; Indels 5; Gaps 4;

QY 3 HLPWTPYIK--LPLNGLEYQPTGLFINNKFVPSKQNTPEVINPSTPEEICHVYEGREDD 60
Db 5 HETAEPEVKITLPGNYTYEPTGLFINNKFVPSKQNTPEVINPSTPEEICHVYEGREDD 64
QY 61 VEEAQAADRAFSNGSWNGIDPIDRGKALRYLAELIEODKDVIASTIETLDNGKATS-SSR 119
Db 65 VEVAIECADRAFHDTEWATQDPRERGRLLSKLADELESQIDLVSSIEALDNGKTLAPKAR 124
QY 120 GDVDLVINLYKSSAGFADKIDGRMIDTGRTHSYTKRQPLGVCGQIIPWNPFLMWAKWI 179
Db 125 GDVTIACLNRDAAYADKVGRTINTDGYMNFITLPIGVCGQIIPWNPFLMWAKWI 184
QY 180 APALVTGNTVLTAEISTPLSALYSKYIPOAGIPGVINIVSGFKIVVEAINTNHPKIK 239
Db 185 APALMGNVCLKPAAVTFLNALYFASICKVGIAGVNVIVPGFHTVGAALTNDDPIR 244
QY 240 KVAFTGSTATGRHI--YQSAAGHKKVITELGKSPNIVFADAEKAKVQNIILGIYVNSG 298
Db 245 KLAFTGSTEVGKSVAVDSSESNLKKITILEGKSAHLVFDGANIKKTLPLNVNGIFKNAG 304
QY 299 EVNAGSRVYVEESIYDKFTIEEPKAASES-IKVGDPDEFSTFGAQTSMQNLKILKYVD 357
Db 305 QICSSGRYIVQEGITDELLAAFKAYLETEIKVGNFDRANFQGAITNRQOFDTIMNYID 364
QY 358 IGKNEGATLITGGERIGSKGYFIKPTVFGDVKRMKVEETFGPVVTVTTRFKSADEVIN 417
Db 365 IGKKEGAKILTGKGVKDGKGYFIRPTVYFVVDNEMRIYKEETFGPVVTVAKFTEEGVE 424
QY 418 MANDSEYGLAAGHTSNINTALKVADRVNAGTVIWTNTNDHNAVFPFGFNASGLGREGMS 477
Db 425 MANSSEFGLSGISTETSLGLKVAKMLKAGTVIWTNTNDFDSRVPFGCVKQSGYGRMG 484
QY 478 VDALQNYLVQKAVRAKL 494
Db 485 EVVHAYTEVAVRIKL 501

RESULT 12
US-60-360-039-3846
Sequence 3846, Application US/60360039
GENERAL INFORMATION:
APPLICANT: Cao, Yonwei
APPLICANT: Chen, Xiantong
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.

Db 485 EVVHAYTEVAVRIKL 500

RESULT 11
US-09-791-537-70619
Sequence 70619, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 70619
LENGTH: 501
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-791-537-70619

Query Match 56.2%; Score 1442.5; DB 27; Length 498;
Best Local Similarity 54.1%; Pred. No. 1.4e-126;
Matches 258; Conservative 88; Mismatches 114; Indels 2; Gaps 4;

QY 25 FINNKFVPSKQNTPEVINPSTPEEICHVYEGREDDVFEAVQAADRAFSNINWNTDID 84
Db 1 FINNEVEGVKKTPEVINPSTPEEICHVYEGREDDVFEAVQAADRAFSNINWNTDID 89
QY 85 PGKALRYLAELIEODKDVIASTIETLDNGKATS-SSR 119
Db 60 RGIYLLKLADELKLLAALAVESLDNGKATS-SSR 119
QY 145 DTGRTHSYTKRQPLGVCGQIIPWNPFLMWAKWIIPWNPFLMWAKWIIPWNPFLMWAKWI 204
Db 120 DISPDSEHYTROEPLGVCGQIIPWNPFLMWAKWIIPWNPFLMWAKWIIPWNPFLMWAKWI 179
QY 295 SKYIPGAGIPGVINIVSGFKIVVEAINTNHPKIKVAEIS-IAI-PPHRYGSA-ATK 264
Db 180 AQVKEAGHPGVLNIIISGFGRIAGAAASHMDIDUKVAFSTWVGKRIKMAAAESNLKK 239
QY 264 VTLEDKSPNIVFADAEKAKVQNIILGIYVNSGFTVTA-SPVYVPEFSTVYFTEPPKA 424
Db 240 VTLEDKSPNIVFADAEKAKVQNIILGIYVNSGFTVTA-SPVYVPEFSTVYFTEPPKA 424
QY 324 ASHSTHVGDPDEFSTFGAQTSMQNLKILKYVDVFNHGALEIDGDEPLSKYVDEKPT 384
Db 300 RAGNVKGVDPDEFSTFGAQTSMQNLKILKYVDVFNHGALEIDGDEPLSKYVDEKPT 399
QY 384 VEGVKEKMKVKEETFGPVVTVTTRFKSADEVINMANISEYGLAAGHTSNINTALKVAD 443
Db 360 IFTNVHEDMKIKKEETFGPVVTVAKFTEEGVE 419
QY 444 RVNAGLVNIIISGFGRIAGAAASHMDIDUKVAFSTWVGKRIKMAAAESNLKK 480
Db 420 HIRAGTVWNTCKYTCFVLEKVTIHRANONI FOOTNALIHQIPRGYKESCHOREDEBA 479
QY 481 LQNYLVQKAVRAKL 495
Db 480 LANYTCKSAVAKLN 494

RESULT 13
US-08-612-853-11
Sequence 11, Application US/08612853
GENERAL INFORMATION:
APPLICANT: Achatz, Gernot
APPLICANT: Oberkofler, Hannes
APPLICANT: Simon, Birgit
APPLICANT: Unger, Andrea
APPLICANT: Lechenauer, Erich
APPLICANT: Hirschwehr, Reinhold
TITLE OF INVENTION: Recombinant Cladosporium Herbarum
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas

APPLICANT: Hirschwehr, Reinhold
 TITLE OF INVENTION: Recombinant Cladosporium Herbarum
 NUMBER OF INVENTIONS: Allergens
 NUMBER OF SEQUENCES: 117
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fendle & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTA Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/612,853
 FILING DATE: 26-FEB-1996
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AT94/00120
 FILING DATE: 02-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Friebe, Thomas E.
 REGISTRATION NUMBER: 29,258
 REFERENCE/LOCKET NUMBER: 6550-021-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 496 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE:
 ORGANISM: Cladosporium herbarum
 DEVELOPMENTAL STAGE: Spores and vegetative hyphae

US-08-612-853-2

Query Match 54.9%; Score 1399; DB 10; Length 496;
 Best Local Similarity 53.8%; Pred. No. 26-122;
 Matches 263; Conservative 89; Mismatches 135; Indels 2; Gaps 2;

QY	7	TVPKILPGLNGLEYHPTGLFIPNNKVPKSKONTFEVINPSTEEETICHIEGREDDOVEAVO	66
DB	3	SVQLETPHSGKYEQPTGLFINNEFVKGGKTFDVINPDSVITQVHEATEKDVDAVA	62
QY	67	AADRAFSNCSWNCIDIPIDRGKALYRIAELEQDKDIVIASIETLNGKRAISSRGEDVLVI	126
DB	63	AAPQAF-PGSGWPIFTFPNFGKILNNLANLEKNTDILLAAVESLDNGKATSMARVTSACAS	121
QY	127	NVILKSSAGFADKIDGRMIDTGRTHFSYTKROPGLWCGGQIPWNEPLLMWANKIAPALVTG	186
DB	122	GCLRYGGWADKLTGKVDTTPTDTNYNKKKEPIGVCRSHLSLELPLLMWANKIGPAIACG	181
QY	187	NTVVLKTAETPLSALYSKYIPQAGIPPGVINIVSGRGIIVFAIINHPKIKKVAFTGS	246
DB	182	NTVVLKTAETPLGGIIVAAASLVKEAGFPFGVINIVISGFKVAGAAALSSHMVDVKVFTGS	241
QY	247	TATGRHIYSAA-AGLKKVLTLELGGKSPNIVFAELKKAQVQNIILGIYNSGEVCCAGS	305
DB	242	TVVGRITLKAASSNLKKVITLELGGKSPNIVFEDADIDNAISWYNFGIFFNHGCCAGS	301
QY	306	RVVVEESIYDKFIEEFKAASESIKVGDPFDESTFOGAQTSOMQLNKILKYVDIGKNEGAT	365
DB	302	RVVQESIYDKFQVKERAKQNVVGDPAALTFQGPVSKVQFDKIMEYIQACKDASAT	361

QY 496 LITGGERLGSKSYETETPIVFGIWKELMKLVKGEIPGPVVIVIKPSADEVINMANISSEYQ 425
 DB 362 VETGSSKSGKSYETETPIESNVIEIMKLVKGEIPGPVSTAKETIKETAIKIGSNASTYQ 421
 QY 426 LAAGIHTSNINIALKVAQHVNAQTVMINLYNDFHVAPEGGENANGSGRENSVIALQNYL 485
 DB 422 LAAAVHTENLTATEVSNALFKATVWVNTYNTIIEUMPPFGYKESGIGFETGFIATANYI 481
 QY 486 QVKAVRAKL 494
 DB 482 QTKTVSIRL 490

Search completed: June 24, 2004, 10:40:46
 Job time : 182.207 secs

GenCore version 5.1.6
Copyright (c) 1994 - 2003 CompuLink Ltd

OM protein protein search, using sw mod-1

Run on: June 24, 2003, 10:14:35 ; Search time 53.87% seconds
(without alignments)
2375.712 Million cell updates/sec

Title: US-09-830-751-2

Perfect score: 2550

Sequence: 1 MSHLPMTVPKLPNGLEYEQ MSVPAIANYLQKAVPAKED 495

Scoring table: BLOS62

Gapop 10.0 ; Gapext 0.5

Searched: 1171708 seqs, 257189365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Pending_Patents_AA_New*

1: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
2: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
3: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
4: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
5: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
6: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*
7: /cgn2_6/prodata/2/paa/paa_NEW_00MK pep.*

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2550	100.0	495	US-09-830-751-2	Sequence 2, Appli
2	2538	99.5	514	US-10-369-493-22438	Sequence 22438, A
3	1728.5	67.8	520	US-10-369-493-1749	Sequence 1749, AF
4	1452	56.9	500	US-10-369-493-22495	Sequence 22495, A
5	1434.5	56.2	498	US-10-369-493-1846	Sequence 846, AP
6	1274.5	50.0	519	US-10-219-051B-6969	Sequence 6969, AP
7	1274.5	50.0	519	US-10-219-051B-6973	Sequence 6973, AP
8	1254.5	49.2	532	US-10-369-493-12695	Sequence 12695, A
9	1253.5	49.2	517	US-10-219-051B-6971	Sequence 6971, AF
10	1253.5	49.2	517	US-10-219-051B-6975	Sequence 6975, AP
11	1253.5	49.1	520	US-09-949-016-10654	Sequence 10654, A
12	1252.5	49.1	495	US-10-369-493-23162	Sequence 3162, A
13	1244.5	48.8	516	US-10-369-493-1555	Sequence 3555, AP
14	1239.5	48.6	514	US-09-453-135-14871	Sequence 14871, A
15	1239.5	48.6	514	US-09-453-135-14871	Sequence 14871, A
16	1239.5	48.6	514	US-09-453-135-14871	Sequence 14871, A
17	1236.5	48.5	544	US-10-369-493-5715	Sequence 5715, AF
18	1234.5	48.4	552	US-10-425-114-59042	Sequence 59042, A
19	1231	48.3	532	US-09-949-016-9297	Sequence 9297, AP
20	1230	48.2	518	US-09-724-676-64510	Sequence 64510, A
21	1230	48.2	518	US-09-724-676-64510	Sequence 64510, A
22	1226.5	48.1	552	US-10-425-114-49963	Sequence 49963, A
23	1225.5	48.1	544	US-10-425-114-51294	Sequence 51294, A
24	1225.5	48.0	500	US-09-830-751-4	Sequence 4, Appli
25	1224	48.0	813	US-10-437-963-11358	Sequence 11358, A
26	1222.5	47.9	518	US-10-144-779-472	Sequence 472, App

ALIGNMENTS

RESULT 1

US-09-830-751-2
Sequence 2, Application US/09830751

GENERAL INFORMATION:

APPLICANT: Suthers, Patrick F

APPLICANT: Suthers, Patrick F

TITLE OF INVENTION: Production of a hydroxyproline A-13 in a recombinant

PATENT REFERENCE: 602496617

CURRENT FILING DATE: 2000-08-30

PRIOR FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/151,440

PRIOR APPLICATION NUMBER: PCT/US00/23878

NUMBER OF SEQ ID NOS: 23

SOFTWARE: Patent Ver 2.1

SEQ ID NO 2

LENGTH: 495

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-830-751-2

Query Match

Best Local Similarity: 100.0%, Score: 2550, DP: 5, Length: 495;

Matches: 495; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY	1	MSHLPMTVPKLPNGLEYEQPTGLFINKKVPKSKONTPEVINSPTPEETICHYEGREDD	60
DB	1	MSHLPMTVPKLPNGLEYEQPTGLFINKKVPKSKONTPEVINSPTPEETICHYEGREDD	60
QY	61	VEFAVQAADPAFSSNNRINDIPKALYPLAELPEHIVIVIASIETLNTKFAISSSPR	120
DB	61	VEFAVQAADPAFSSNNRINDIPKALYPLAELPEHIVIVIASIETLNTKFAISSSPR	120
QY	121	LVTLVINYKSSAFKRIKPMIDGKTHFSTYKPKPLDVGQIIPWNEPLLMWAKIA	180
DB	121	LVTLVINYKSSAFKRIKPMIDGKTHFSTYKPKPLDVGQIIPWNEPLLMWAKIA	180
QY	181	PALVGTNIVLTKAESTPI;SALVYSKYTPGAGIPPEVINIVSGFGKIVVATINHPKIKK	240
DB	181	PALVGTNIVLTKAESTPI;SALVYSKYTPGAGIPPEVINIVSGFGKIVVATINHPKIKK	240
QY	241	VAFGSTATDRIYQSAAGLKFVLELEGGSKSNIVAIALPYAVNIIILGIIYNSGEV	300
DB	241	VAFGSTATDRIYQSAAGLKFVLELEGGSKSNIVAIALPYAVNIIILGIIYNSGEV	300
QY	301	CAVSSFFVVEESIVYKPIPEKPAASSTIKVTDIPESTPEAVTSAMQINPKTIKYVDTCP	360
DB	301	CAVSSFFVVEESIVYKPIPEKPAASSTIKVTDIPESTPEAVTSAMQINPKTIKYVDTCP	360

Db 301 CCAGSRVYVEES IYDKFIEEFKAASESIKVGDPDFDSTFGAQT SOMQLNKILKYVDICK 360
QY 361 NEGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 420
Db 361 NEGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 420
QY 421 DSEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 480
Db 421 DSEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 480
QY 481 LONYLOVKAVRAKLD 495
Db 481 LONYLOVKAVRAKLD 495

RESULT 2
US-10-369-493-22438
: Sequence 22438, Application US/10369493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 22438
: LENGTH: 519
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22438

Query Match 99.8%; Score 2538; DB 6; Length 519;
Best Local Similarity 99.8%; Pred. No. 2, 1e 196;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 SHLPMTVPILKPLNGLEVEQPTGLFINNKVPVSKQNTFEVINPSTEEFICHIEGREDV 61
Db 25 SHLPMTVPILKPLNGLEVEQPTGLFINNKVPVSKQNTFEVINPSTEEFICHIEGREDV 84
QY 62 BEAVQAADRAFNSWNGIDPDRGKALYPLAELIEQDKDVIASITELDKGATSSRGD 121
Db 85 BEAVQAADRAFNSWNGIDPDRGKALYPLAELIEQDKDVIASITELDKGATSSRGD 144
QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPPLLMWAKIAP 181
Db 145 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPPLLMWAKIAP 204
QY 182 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEAINTNHKIKKV 241
Db 205 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEAINTNHKIKKV 264
QY 242 AFTGSTATGRHIYQSAAGLKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 301
Db 265 AFTGSTATGRHIYQSAAGLKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 324
QY 302 CAGSRVYVEES IYDKFIEEFKAASESIKVGDPDFDSTFGAQT SOMQLNKILKYVDIGKN 361
Db 325 CAGSRVYVEES IYDKFIEEFKAASESIKVGDPDFDSTFGAQT SOMQLNKILKYVDIGKN 384
QY 362 EGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 421
Db 385 EGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 444
QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 481
Db 445 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 504

QY 482 ONYLOVKAVRAKLD 495
Db 505 ONYLOVKAVRAKLD 518

RESULT 3
US-10-369-493-1749
: Sequence 1749, Application US/10369493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: CURRENT FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 1749
: LENGTH: 520
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1749

Query Match 67.8%; Score 1228.5; DB 6; Length 520;
Best Local Similarity 64.8%; Pred. No. 5, 1e 141;
Matches 320; Conservative 81; Mismatches 92; Indels 1; Gaps 1;

QY 2 SHLPMTVPILKPLNGLEVEQPTGLFINNKVPVSKQNTFEVINPSTEEFICHIEGREDV 61
Db 24 SOAPLAVPITLPGCTEYEQPTGLFINNVPVSKQNTFEVINPSTNEEKITIVYKAMEIDV 84
QY 62 BEAVQAADRAFNSWNGIDPDRGKALYPLAELIEQDKDVIASITELDKGATSSRGD 121
Db 84 BEAVQAADRAFNSWNGIDPDRGKALYPLAELIEQDKDVIASITELDKGATSSRGD 142
QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPPLLMWAKIAP 181
Db 143 VALVSKYLRSGGWAIRKIVGNVITDGNHIFYSIKPELVAGQIIPWNPPLLMWAKIAP 202
QY 182 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEAINTNHKIKKV 241
Db 203 ALAIGNIVVIVKVAETPLSALFASQIQKAGIPAGVNIIPGSRKVVGERLSABHVKKI 262
QY 242 AFTGSTATGRHIYQSAAGLKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 301
Db 263 AFTGSTATGRHIMKVAANDIVKVTLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 322
QY 302 CAGSRVYVEES IYDKFIEEFKAASESIKVGDPDFDSTFGAQT SOMQLNKILKYVDIGKN 361
Db 323 CAGSRVYVEES IYDKFIEEFKAASESIKVGDPDFDSTFGAQT SOMQLNKILKYVDIGKN 382
QY 362 EGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 421
Db 383 EGATLITGGERHASKGYFIKPTVFGVKEDMPIVKEEIPGPVVITKPKSADEVINMAN 442
QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 481
Db 443 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDHFHVAVPFGGFNAGSLGREGMSVDA 502
QY 482 ONYLOVKAVRAKLD 495
Db 504 SNTYOTKRSVRIAD 516

RESULT 4
US-10-369-493-22495
: Sequence 22495, Application US/10369493


```

; SOFTWARE: Perl script
; SEQ ID NO 6969
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
; DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-6969

```

```

Query Match      50.0%   Score 1274.5, DB 6, Length 519,
Best Local Similarity 50.8%   Pred. No. 2 8e 94;
Matches 252; Conservative 85; Mismatches 146; Indels 13; Gaps 6;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKFPVPSKONKTFEIVINPSTEEICHIEGREDDE 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24 TSAVPAPN-----QOPEVFCNQIFINNEWHDAVSKKTFEPTVNPSTGEVICOVAEGKEDVD 79
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 EAVQAADRAFNSGS--WNGIDPIDRGKALYRLAELIADODKDVIASTETLDNGKA-ISSSKG 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 80 KAVKAAQAALQGLSPWRMDASDRGLLYRLADLIERDRTYLAALETLDNGKPYVISYLV 139
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 DVDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQIIPWNPFLMLMAWKIA 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 140 DLDVVLKCLRYAYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCQIIPWNPFLMQAWKLG 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 PALVTGNTVYLVKTAESTPLSALYSKY IPOAGIPPGVINIVSGFGKIIVVEAITNHPKTK 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 PALATGNVVMKVAEQPTLALYVANLKEAGFPVGVNIVPGPGPTAGALASHEDVDK 259
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 VAFPGSTATGRHLYQSAA--AGLKVYTLLELGGKSNIVFADELKKAUVNIILGYYNSG 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 VAFPGSTEVG--HLIQVAAGSSNLKRVTLLELGGKSNILMSDADMDWAVEQAHFALFFNQ 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 299 EVCCAGSRVYVESIYDKFTEERKAASESIKVDPPDESTFGAQTQSOMLNKILKYVDI 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 OCCAGSRTEVQEDVDYDEFVRSVARAKSRVGNPNDSRTEGQGVDETFQPKILLYIKS 378
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 359 GKNEGATLITGGERLGSKGFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADDEVINM 418
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 GQEGAKLCCGGAADAAFFYIQPTVFGDVKDGMTIAKEEIPGVNQILKFKTIEEVGR 438
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 419 ANDSEYGLAAGIHTSNITALKVADRVNACTVWINTYNDPHAVPFGGFGNAGLGREMSV 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 ANSKYGLAAAVFTKLDKANYLSQALQAGTVMINCYDVEGAQSPFGYKMGSGSGRELGE 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 479 DALONYLQVKAARAKL 494
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 499 YGLQAYTEKTVTKV 514
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-10-219-051B-6973
; Sequence 6973, Application us/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Lea 35693 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/219,051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715
; SOFTWARE: Perl script
; SEQ ID NO 6973
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Rattus norvegicus

```

```

; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
; DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-6973

```

```

Query Match      50.0%   Score 1274.5, DB 6, Length 519;
Best Local Similarity 50.8%   Pred. No. 2 8e 94;
Matches 252; Conservative 85; Mismatches 146; Indels 13; Gaps 6;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKFPVPSKONKTFEIVINPSTEEICHIEGREDDE 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24 TSAVPAPN-----QOPEVFCNQIFINNEWHDAVSKKTFEPTVNPSTGEVICOVAEGKEDVD 79
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 EAVQAADRAFNSGS--WNGIDPIDRGKALYRLAELIADODKDVIASTETLDNGKA-ISSSKG 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 80 KAVKAAQAALQGLSPWRMDASDRGLLYRLADLIERDRTYLAALETLDNGKPYVISYLV 139
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 DVDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQIIPWNPFLMLMAWKIA 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 140 DLDVVLKCLRYAYAGWADKYHGKTIPIDGDFFSYTRHEPVGVCQIIPWNPFLMQAWKLG 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 PALVTGNTVYLVKTAESTPLSALYSKY IPOAGIPPGVINIVSGFGKIIVVEAITNHPKTK 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 PALATGNVVMKVAEQPTLALYVANLKEAGFPVGVNIVPGPGPTAGALASHEDVDK 259
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 VAFPGSTATGRHLYQSAA--AGLKVYTLLELGGKSNIVFADELKKAUVNIILGYYNSG 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 VAFPGSTEVG--HLIQVAAGSSNLKRVTLLELGGKSNILMSDADMDWAVEQAHFALFFNQ 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 299 EVCCAGSRVYVESIYDKFTEERKAASESIKVDPPDESTFGAQTQSOMLNKILKYVDI 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 OCCAGSRTEVQEDVDYDEFVRSVARAKSRVGNPNDSRTEGQGVDETFQPKILLYIKS 378
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 359 GKNEGATLITGGERLGSKGFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADDEVINM 418
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 GQEGAKLCCGGAADAAFFYIQPTVFGDVKDGMTIAKEEIPGVNQILKFKTIEEVGR 438
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 419 ANDSEYGLAAGIHTSNITALKVADRVNACTVWINTYNDPHAVPFGGFGNAGLGREMSV 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 ANSKYGLAAAVFTKLDKANYLSQALQAGTVMINCYDVEGAQSPFGYKMGSGSGRELGE 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 479 DALONYLQVKAARAKL 494
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 499 YGLQAYTEKTVTKV 514
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 8
US-10-369-493-12695
; Sequence 12695, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52053)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12695
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(532)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12695

```


Query Match: 49.2%, Score 1254.5, DB 6, Length 517,
 Best Local Similarity 47.0%, Pred. No. 1.4e-92;
 Matches 248; Conservative 88; Mismatches 151; Indels 41; Gaps 5;

QY 8 VPIKLPNGLEYEQPTGLFINNKVPSPKQNTFEVNP-----STEEDE 49
 Db 1 VSLTAPNGHKYEQPTGLFINNFVASKSGEKFATVNPQVCSGDARXXLYSNIRASDEEE 60

QY 50 ICHIVEPFCVFAVAAAFASNSJWNIIPKSKALYFLAELIEQKVIASIFIL 109
 Db 61 ITQVAGEEDIDIAVAAKALKDPSKLLTIDRGNLMLKLADLIDONKETLAVETW 120

QY 110 DNE-----KATIS-----SSPDEVCLVINVLKSSAFALKIICPMIITGPTHF 151
 Db 121 LWXNLISLSQPIILXLSXAKPYQVSLNFIISVVNIIPYCAQWALKIHQOTISTTPAKF 180

QY 152 SYTKKQPIGVGCGQIIPNFFLLMAMWIAFALVTGNTVVLKTAESTLSALYVSKYFOA 211
 Db 181 AYTLRQPIGVVQIIPNFFPLMAAKKLSPALAGSNTVVLKPAEQIPLSLILYLAKFIKEA 240

QY 212 GIFTGVINIVSCFCKIVVEATTNFKTKKVAFTGCTATGPHIYQSAAGIKKVTILGCK 271
 Db 241 GPPCVVNVNGLGRVACSLVTHPGVDKVAFTGCTMTGKRTMKMAAGTKNNVTDETCK 300

QY 272 SPNIVFADAEKKAQVNIILGIYNSGEVCCAGSRVYVEESYIDKFIIEBFK-AASESIKV 330
 Db 301 SPILVFDADIPOAKWAHIGIMYNOGQVCTATSPILVHEKVDHFFILPREAVATTSKV 360

QY 331 GPHFESTIFQAQISQMLNKLILKYVDIGKNDJALILRG-----ERLKGKGYFKPTVPG 386
 Db 361 GUPESDDIFUGQVTKAUYERVLVSYLESGKQESATLVGGVGPYKNVKGKGFPIAPTIFT 420

QY 387 DVKEDMRIVKEIFGPPVTVTKPKSADEVINMANDSYGLAAGHTSINTALKVADRYN 446
 Db 421 NVKNNPIRYFVFGPPVATAPFSTEEAIPRANVTYGLCAAVFTKDIERAKRVASEIE 480

QY 447 AGTVWINTYNDHFHAVPPGFNAGSLGRFMSVDALQNYIQVKAVRAKL 494
 Db 481 AGMWINSNSLSEFVFGVVKVSGIGFELGEALSLAYIQLKAVHVM 548

RESULT 9
 US-10-219-051B-6971
 ; Sequence 6971, Application US/10219051B
 ; GENERAL INFORMATION:
 ; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 ; APPLICANT: Hospital / Bayer AG
 ; TITLE OF INVENTION: Nucleotide sequences involved in pain
 ; FILE REFERENCE: Lea 35693 Foreign Countries
 ; CURRENT FILING DATE: 2003-05-09
 ; PRIOR FILING DATE: 2001-08-14
 ; PRIOR FILING DATE: 2001-11-01
 ; PRIOR FILING DATE: 2001-11-01
 ; PRIOR FILING DATE: 2001-11-01
 ; NUMBER OF SEQ ID NOS: 14715
 ; SOFTWARE: Perl script
 ; SEQ ID NO 6971
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 ; DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6971

Query Match: 49.2%, Score 1253.5, DB 6, Length 517,
 Best Local Similarity 50.3%, Pred. No. 1.4e-92;
 Matches 249; Conservative 80; Mismatches 149; Indels 11; Gaps 5;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKVFSPKQNTFEVNPSTFEICHIEGREDUVE 62

Db 22 IOAVFAIN- QQVFEVCNUIFINNEWHUAVSRKTFPTVNFSTGEVICQVAGKEEVD 77

QY 63 EAVUADAKFASNGS-WMGIDPIDRGKALYLAELIEQDKDVIASIFELANGKA-ISSSRG 120
 Db 78 KAVKAAFAAFULASPPMPFASHPGLINPLAULLEGPYLAALIEILNGSKPYVVISLV 137

QY 121 DVDLVINYLKSSAGFADKIDGPMIDGPTHFSTKPOPLGVCQGIIPWNFPPLIMAAWKTA 180
 Db 138 DLNWLKCTRYAGWAKNYHGKTIPTDGDFFSTPTPEHPGVSCQIIPWNFPPLIMAAWKLG 197

QY 181 FALVTGNTVVLKTAESTLSALYVSKYIPQAGIFFGVINIVSOFKPVVVEAITHPIKPK 240
 Db 198 PALATGHWVMKVAFTFETLALYVNLKFAEGFFPFWNI-VFQVGTACAAIASIHEDVKE 257

QY 241 VAFITGCTATGPHIYOSA-AAGLKKVILELGKSNIVFALAEKKAQVNIILGIYNSGE 299
 Db 258 VAFITGCTATGPHIYOSA-AAGLKKVILELGKSNIVFALAEKKAQVNIILGIYNSGE 317

QY 300 VCCAGSRVYVEESYIDKFIIEEKAASESIKVGPPDEFTFGAQTSSOMOLNKLILKYVDIG 359
 Db 318 CCCAGSPTFVQERIYREFVRSVARPAKSPVVGPNPFTSKTEGGQVDETQPKKILKYINIG 377

QY 360 KNEKATILITSEERIGSKGYFKPTFGDVKEDMRIVKEIFGPPVTVTKPKSADEVINMA 419
 Db 378 KQFCAKILITGEGGIAADPSEYFIQPTVFGDVQDGMTIAKEEIEFGPMILKPKTIFEVVGRA 437

QY 420 NDSEYGLAAGHTSINTALKVADRYNAGTWTWINTYNDHFHAVPPGFNAGSLGRFMSVD 479
 Db 438 NNSTYGLAAAVFTKLEKANYLSQALQAGTVWVNCYVDVFGAQPFGYKMSGSGRELGRY 497

QY 480 ALQNYLQVKAVRAKL 494
 Db 498 GLQATTEVKTIVTKV 512

RESULT 10
 US-10-219-051B-6975
 ; Sequence 6975, Application US/10219051B
 ; GENERAL INFORMATION:
 ; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 ; APPLICANT: Hospital / Bayer AG
 ; TITLE OF INVENTION: Nucleotide sequences involved in pain
 ; FILE REFERENCE: Lea 35693 Foreign Countries
 ; CURRENT FILING DATE: 2003-05-09
 ; PRIOR FILING DATE: 2003-05-09
 ; PRIOR FILING DATE: 2001-08-14
 ; PRIOR FILING DATE: 2001-11-01
 ; PRIOR FILING DATE: 2001-11-01
 ; PRIOR FILING DATE: 2001-11-01
 ; NUMBER OF SEQ ID NOS: 14715
 ; SOFTWARE: Perl script
 ; SEQ ID NO 6975
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 ; DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6975

Query Match: 49.2%, Score 1253.5, DB 6, Length 517,
 Best Local Similarity 50.3%, Pred. No. 1.4e-92;
 Matches 249; Conservative 86; Mismatches 149; Indels 11; Gaps 5;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKVFSPKQNTFEVNPSTFEICHIEGREDUVE 62

Db 22 IOAVFAIN- QQVFEVCNUIFINNEWHUAVSRKTFPTVNFSTGEVICQVAGKEEVD 77

QY 63 EAVUADAKFASNGS-WMGIDPIDRGKALYLAELIEQDKDVIASIFELANGKA-ISSSRG 120

QY 360 KNEGATLITGGERLSSKGYFTKPIVGLWEDMFIKKEIFGPVVIVTKPKSADEVINMA 419
 Db 359 KDEGAKAVTGGSCPFAGYFAPVFANVEDMTIAKEEIFGVLTAIPYETVDEVIERA 418
 QY 420 NDEYGLAAGIHTSNTINTALKVADRVNAGTVWINTYNDFIHIAVPEGFGNASGLGREGMSVD 479
 Db 419 NHSEYGLAAGLWENVKQAHYIADRLQAGTVWVCYNVFDAAAPFGYKOSGLGREGMSY 478
 QY 480 ALQNYLOVKAVRAKID 495
 Db 479 ALDNYTEVKSVMNLE 494
 RESULT 13
 US-10-369-493-3555
 ; Sequence 3555, Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Harry S.
 ; APPLICANT: Chen, Xiaofeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10/520521B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR APPLICATION NUMBER: US 09/400,049
 ; PRIORITY FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 3555
 ; LENGTH: 516
 ; TYPE: PRT
 ; ORGANISM: Neurospora crassa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(516)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-3555

Query Match 48.8%; Score 1243.5; DB 6; Length 516;
 Best Local Similarity 47.3%; Pred. No. 8.7e-92;
 Matches 342; Conservative 92; Mismatches 155; Indels 23; Gaps 4;
 QY 6 MTVPILKPLNGLEYEPTGLFINNFVPSKQNKTFEVIN-----PST 46
 Db 5 MEVELTAPNGKKWQPLGLFINNEFVSANEQKLISINPTVYVSNPISLIXLRVLTIPST 64
 QY 47 EEEICHIVYEGREDDVEEAQVADRAFNGSNGNGIDPIDRGKALYRIAELEQDKQVIASI 106
 Db 65 EEFICSVYAATAEDVDAVSAARAFRHSWKSLSGTGERGALMRKPLADLVAENAEILATI 124
 QY 107 ETLDNCKALSSRGD-VDLVINYIKSSAGFADKIPDEMILTGKTHFSYTKPQFLGVCQOI 165
 Db 125 ECLDNKGYQVQALNENPVEINLVRYAGYADKNFGQVIDVGPAKFAITVKEPLGVCQOI 184
 QY 166 IPWNPPLLMWAKIAPALVTGNTVLTAEPTLSALYVSKYIPQAGIPPGVINIVSGFG 225
 Db 185 IPWNPPLMAAKLGPALCGNTVVLKLAQVPLSVLYLAKLKEAGFPVGVNLINGH 244
 QY 226 KIVVEATINHPKIKKVAFTGTATGRHIIYOSAAAGLKKVTLGCGKSPNIVFADAEIKKA 285
 Db 245 RPAGAAIVQHPQVTKIAPTGTGTEIKMKMASYTMKNITILETGKSPITVPEADALELA 304
 QY 286 VQNIILGIYNSGWCAGSVVVEESLYDKFTEERKKAASESIKV-GDPEDESTFOGAOT 344
 Db 305 ATWSHIGIMNSGOQICATISRLVBEKIYDEFVEKFAKVOEVSVLGDPEESTFHPQV 364
 QY 345 SUMQLNKILKYVDIGKNNGATLITGGEFL--GSGNGYFIKPTVFSUWKEDMPVKEEIFGP 402
 Db 365 TKAQYERVLGVINVKRPGATVMGTSEPAFQNSKGFVAPIVETNVKPTMKIIFPEEIPGP 424

QY 403 VVIVTKPKSALEVINMANUSYGLAASTHTSNTINTALKVADRVNAGTVWINTYNDFIHIAV 462
 Db 425 QVALITPKTEBEALTLANDSMYSGAALPTFTGTRAHFPVAPFTEAGVWVWVSSNDSDFRJ 484
 QY 463 PFGGPNASGLGKREMSVDALQNYLOVKAVRAKL 494
 Db 485 PFGGVKQSNIGFELGEACLAPYCNVKSITHVNL 516
 RESULT 14
 US-60-453-135-14871
 ; Sequence 14871, Application US/60453135
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele
 ; APPLICANT: IAKOUBOVA, Olga
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; FILE REFERENCE: MYOARADIAL INHIBITION, MECHANISMS OF DEFENTION AND USES THEREOF
 ; CURRENT APPLICATION NUMBER: US/60/453,135
 ; CURRENT FILING DATE: 2003-03-10
 ; NUMBER OF SEQ ID NOS: 82762
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14871
 ; LENGTH: 514
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-453-135-14871
 Query Match 48.6%; Score 1239.5; DB 7; Length 514;
 Best Local Similarity 50.4%; Pred. No. 1.8e-91;
 Matches 246; Conservative 85; Mismatches 146; Indels 11; Gaps 5;
 QY 7 TVPIKLPLNGLEYEQP----TGLFINNFVPSKQNKTFEVINPSTEEIECHIVYEGREDDVE 62
 Db 22 TQAVFAPN----QQFVFPQNOIFINNEHDAVSKRTFETVNPSTGEVICOVAGDEKDDVD 77
 QY 63 EAVGAADPAFNGS-WNGIPIPIRGKALYPIAELEQDKQVIASITETIDNGKA-ISSSPG 120
 Db 78 KAVKAARAAAFOLGSPWPPMDASHPGPLNLRLADLIERDRTYLALETLDNGKPYVIVSYLV 137
 QY 121 DVDLVINYIKSSAGFADKIPGMDTGTTHFSYTKPQFLGVCQOIIPWNPPLLMWAKIA 180
 Db 138 DLDVNLKRLRYAGWADQYHGTIPIGIDFESTHPEPVGVCQIIPWNPPLLMWAKLG 197
 QY 181 PALVTGNTVLTAEPTLSALYVSKYIPQAGIPPGVINIVSGFGKIVVEATINHPKIKK 240
 Db 198 PALAIGNVWVWVKVAEQIPLIALLYVANIKEAGFPVGVNLIVPGFPIAGAAIASHEDVDK 257
 QY 241 VAPTGTATRHITYCSA-AAGLKKVTLGCGKSPNIVFADAEIKKAVQNIILGIYNSGF 294
 Db 258 VAPTGTSTEDGVILQVAAVSSNLKPKVTLGCGKSPNILMSALMLWAVEGAHFAHFAH 317
 QY 306 VCGASRKYVVEESLYDKFIEEFKKAASESIKVGEPEDESIFGAGCTSQMLNKLILKYVDIG 359
 Db 318 QVCGASRKYVVEESLYDKFIEEFKKAASESIKVGEPEDESIFGAGCTSQMLNKLILKYVDIG 377
 QY 360 KNEGATLITGGERLSSKGYFTKPIVGLWEDMFIKKEIFGPVVIVTKPKSADEVINMA 419
 Db 378 KQECAKTITGGGTAAGAGCYEPLQIVPGLVQVQSMILAKFPFGDVMILKFKTILEEVVGRA 437
 QY 420 NDEYGLAAGIHTSNTINTALKVADRVNAGTVWINTYNDFIHIAVPEGFGNASGLGREGMSVD 479
 Db 438 NNSYVLAAGVYETKRIKANYISAGLQAGLWVWVYHAPVTAQSEFQYKMSGNSGREGI 497
 QY 480 ALQNYLOV 487
 Db 498 GLQAYTEV 505
 RESULT 15
 US-60-453-050-14871
 ; Sequence 14871, Application US/60453050
 ; GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: SIENOSIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT REFERENCE: C0001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-14871

Query Match.      48.68; Score 1239.5; DB 7; Length 514;
Best Local Similarity 50.48; Pred. No. 1.8e-91;
Matches 246; Conservative 85; Mismatches 146; Indels 11; Gaps 5;

QY 7 TVPIKLPNGLEYEP---TGLFINNKVPFSQNKTFEIVNPSTREEICHIVEGREDDVE 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 22 TOAVPAEN ...QQPEVCNQIFINNEHDAVSRKTFPTVNESTGEVICQVAEGDKREDVD 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 EAVOAAADRAFSNGS-WNGIDIPIDRGKALYRLAELIFODKDVIASTETLDNCKA-ISSSRG 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 KAVKAARAAFOLGSPWRMRDASHRGRLNKLADLIERORTYLAALETLDNGKPYVISLV 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 DVDLVINYLKSSAGPADKIDGRMDIGRTHFSYTKRQPLGVCQGIIPNFPLLMHWAKIA 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 DLDVLKLCRYAGWADKYHGKTIPIIDGDFSYTRHPVGCGQIIPNFPLLMHWAKLG 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGK1VVEAITNHPKIKK 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 PALATGNVVMKVAEQOTPLTALYVANLKEAGFPQGVNIVPGFPTAGAAIASHEDYDK 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 VAFGTGSTATGRHIYQSA-AAGLKVYTLLELGGKSPNIVFADAEKKAQVNIILGIYNSGE 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 VAFGTGSTEIGRVTQVAAGSSNLKRVTLLELGGKSPN1MSDADMDUWAVEQAHFALFFNOGQ 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 VCCAGSRVYVEESIYDKFTEEFKAASESTKVGDPDESTFOGAGTSMOLNKLKLYVDIG 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 CCCAGSRTFVORDIYDEFVERSVARAKSRVGNPFDSKIEQQQVDETFQFKILGYINTG 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 KNEGATLITGGELGSKGYFIKPTVFGVVKEDMPIVKEEIPGVVVTITKPKSADEVINMA 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 378 KQEGAKILCGGIIAADRGVFIQPTVFGVQDCMTIAKEEIPGVVQILKPKTIEEVVGRA 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 420 NDSEYGLAAGIHTSNINTALKYADRVNAGTVMINTYNDFHHAVPFGGENASGLGKREMSVD 479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 438 NNSTYGLAAVFTKDLDKANYLSQALQAGTVWVNCYDVFAGOSPFGGYKMSGSGRELGEY 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 ALQNYIQV 487
   : : : : :
Db 498 GLQAYTEV 505
```

Search completed: June 24, 2003, 10:35:19
Job time : 54.5876 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 24, 2003, 10:10:05 : Search time 18 9426 Seconds
(without alignments)
2512.144 Million cell updates/sec

Title: us-09-830-751-2
Perfect score: 2550
Sequence: 1 MSHLPMTVPKLPNCLEVEQ MSVDALGNYLQKAVPAKLD 435

Scoring table: BLOSUM62
Gap (16 0) , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match %	Length	DR	ID	Description
1	2538	99.5	519	1	S67286	probable aldehyde
2	1728.5	67.8	520	1	S05076	probable aldehyde
3	1452	56.9	500	1	S69929	probable aldehyde
4	1412.5	55.4	497	1	A29055	aldehyde dehydroge
5	1399	54.9	496	2	S43114	aldehyde dehydroge
6	1362.5	53.4	495	2	S43108	aldehyde dehydroge
7	1362	53.4	503	2	T39216	aldehyde dehydroge
8	1275.5	50.0	519	1	I48966	aldehyde dehydroge
9	1274.5	50.0	519	1	S03564	aldehyde dehydroge
10	1253.5	49.2	517	1	DEHUE2	aldehyde dehydroge
11	1252.5	49.1	445	2	H69614	aldehyde dehydroge
12	1252.5	49.1	517	1	A40872	aldehyde dehydroge
13	1240.5	48.6	520	2	S06040	aldehyde dehydroge
14	1236	48.5	544	2	D88449	protein F54D8.3 [i
15	1232.5	48.3	549	2	T03983	rf2 nuclear restor
16	1222.5	47.9	499	2	S74224	aldehyde dehydroge
17	1220.5	47.9	500	1	S00364	aldehyde dehydroge
18	1218.5	47.8	512	1	A35684	aldehyde dehydroge
19	1216.5	47.7	542	2	T02301	aldehyde dehydroge
20	1212.5	47.5	509	1	S14639	aldehyde dehydroge
21	1205.5	47.3	498	2	C83717	NADP-dependent ald
22	1201	47.1	538	2	T06683	aldehyde dehydroge
23	1192.5	46.8	497	2	T64924	aldehyde dehydroge
24	1189	46.6	519	2	C86372	hypothetical prote
25	1182.5	46.4	501	1	DEHFE1	aldehyde dehydroge
26	1181.5	46.3	501	2	S14752	aldehyde dehydroge
27	1179.5	46.3	501	2	J05553	aldehyde dehydroge
28	1176.5	46.1	500	1	S02302	aldehyde dehydroge
29	1172.5	46.0	496	2	T06772	probable aldehyde

30 1170.5 45.9 501 1 I61004 aldehyde dehydroge
31 1169.5 45.9 501 1 A32616 aldehyde dehydroge
32 1168.5 45.8 501 2 J04524 aldehyde dehydroge
33 1141 44.7 498 1 S43184 aldehyde dehydroge
34 1101.5 43.2 496 1 A46725 omega-crystallin -
35 1084 42.5 494 2 B95411 probable aldehyde
36 1081 42.4 506 1 S54615 aldehyde dehydroge
37 1057 41.5 506 1 S54527 aldehyde dehydroge
38 1034.5 40.6 490 2 A66696 glycine betaine al
39 1020 40.0 495 2 C89778 hypothetical prote
40 1006 39.5 511 1 S41308 aldehyde dehydroge
41 1002.5 39.3 500 2 A04544 betaine aldehyde d
42 976 39.1 496 2 C87638 aldehyde dehydroge
43 973 38.9 492 2 A60560 formyltetrahydrofo
44 992 38.9 495 2 H83136 probable aldehyde
45 976.5 38.3 507 2 A03187 aldehyde dehydroge

ALIGNMENTS

RESULT 1

S67286

probable aldehyde dehydrogenase (NAD) (Pir 1 2 1 3) YOR374W - yeast (Saccharomyces cer
N: Alternate names: hypothetical protein U6730
C: Species: Saccharomyces cerevisiae
C: Date: 12-Jul-1996 #Sequence_revision 11-Jan-1997 #text_change 01-Jun-2002
C: Accession: S67286

R: Delius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A: Reference number: S67261

A: Accession: S67286

A: Molecule type: DNA

A: Residues: 1-519

A: Cross references: EMBL:Z75282; PIDN CAA99705.1; GSPDR GNM0015; MIPS-YOR374W

A: Experimental source: strain S288C

C: Genetics:

A: Gene: SGD:ALD7; MIPS:YOR374W

A: Cross references: SGD:S0005901; MIPS:YOR374W

A: Map position: 15R

C: Function:

A: Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A: Note: enzymes with this activity are involved in diverse metabolic pathways in vari
C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C: Keywords: NAD, oxidoreductase
E: 81-344/Domain: aldehyde dehydrogenase homology
F: 290,324/Active site: Glu, Cys #status predicted

Query Match 99.5%; Score 2538; DB 1; Length 519;
Best Local Similarity 99.8%; Pred. No. 4.9e-159;
Matches 493, Conservative 0, Mismatches 1, Indels 0; Gaps 0;

Q7 2 SLPMTVPKLPNCLEVEQPTGLFNNKVFVSKNKTFFVNISSIEFHCHVCEHUV 61
|||||
Db 25 SLPMTVPKLPNCLEVEQPTGLFNNKVFVSKNKTFFVNISSIEFHCHVCEHUV 84
|||||
Q7 62 EFAVAAAPAFSNRWNTETITFGKALYPLAELEGGKIVASIEHINAEZAISSSGD 121
|||||
Db 85 EFAVAAAPAFSNRWNTETITFGKALYPLAELEGGKIVASIEHINAEZAISSSGD 144
|||||
Q7 122 VDLVINLYKSSAFACKTDEPMTHDGFTHSYKKKFGVAVGQIIPNFPPLMAMKAP 181
|||||
Db 145 VDLVINLYKSSAFACKTDEPMTHDGFTHSYKKKFGVAVGQIIPNFPPLMAMKAP 204
|||||
Q7 182 ALVTGNTVWLTAETIFLSALYSKYTPQAGIFGVINIVSGKIVGKIVGKIVGKIV 241
|||||
Db 245 ALVTGNTVWLTAETIFLSALYSKYTPQAGIFGVINIVSGKIVGKIVGKIVGKIV 264
|||||
Q7 242 AFTGTATGRHHYSSAAAGKKEVTFGGKSPNVFAACAFVAVCNHCHVYNSDEV 301
|||||
Db 265 AFTGTATGRHHYSSAAAGKKEVTFGGKSPNVFAACAFVAVCNHCHVYNSDEV 324
|||||
Q7 302 GAGSPVYVESLYKFTTEEFKAASESTKVGTHFEFTFPAQLQAMOLRIKLYVLGKN 361
|||||

Db 325 CAGSRVTVESIVDKFLEEFKAASESIKVGDPDESTFOGAOTSOMLNKILKYVDIGKN 384
 QY 362 EGATLITGGRIAGSKGYFIKPTVFGDVQKEDMRIVKKEIFGPVTVTKFKSADEVINMAND 421
 Db 385 EGATLITGGRIAGSKGYFIKPTVFGDVQKEDMRIVKKEIFGPVTVTKFKSADEVINMAND 444
 QY 422 SEYGLAAGIHTSNITALKVADRVNAGTVMINTYNDPHFHAVPFGGPNASGLGRENMSVDAL 481
 Db 445 SEYGLAAGIHTSNITALKVADRVNAGTVMINTYNDPHFHAVPFGGPNASGLGRENMSVDAL 504
 QY 482 QNYLOQKAVRAKLD 495
 Db 505 QNYLOQKAVRAKLD 518

RESULT 2

S50576

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) YER073w yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C>Date: 28 May 1993 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002

C:Accession: S50576

R:Blotrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: The sequence of S. cerevisiae lambda clone 3612 and cosmid 9747.

A:Reference number: S50438

A:Accession: S50576

A:Molecule type: DNA

A:Residues: 1-520 <DLE>

A:Cross references: EMBL:U08814; NID:G60400; PDB:1AAR64612 1; PDB:9G0410; GSPDB:G00000

C:Experimental source: strain S288C (AB972)

C:Genetics:

A:Gene: SGD:ALD5; MIPS:YER073w

A:Cross references: SGD:S0000875

A:Map position: 5P

C:Function:

A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water

A>Note: enzymes with this activity are involved in diverse metabolic pathways in various

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism, NAD, oxidoreductase

F:82-342/Domain: aldehyde dehydrogenase homology <ALDD>

F:288,322/Active site: Glu, Cys #status predicted

Query Match 67.8%; Score 1728.5; DB 1; Length 520;

Best Local Similarity 64.8%; Pred. No. 7.2e-106;

Matches 320; Conservative 81; Mismatches 92; Indels 1; Gaps 1;

QY 2 SHLPMTVPILKPLNLEWEPTGLFINNKFVPSKQNTFEVINPSTFEFTCHYEGPDV 61

Db 24 SQAPLRVITLPGFTYEPTGLFNGEFVASKQKRTFDVINPSNEEKITTVYKAMEDDV 83

QY 62 BEAVQAAADRAFSGNSWNGIDPIDRQKALYRLAFILPQKQDVIASIETLQNGKATSSSGD 121

Db 84 DEAVAAAKAFET-KWSIVEPEVRKALNPLADLVEKHQETLAATESMUNGSLPCASGD 142

QY 122 VDLVINYLKSSAGFAUKIDGRMIDTGRTHSYTKRQPGVGGQIIPWNFFLLMNAWKIA 181

Db 143 VALVSKYLRSCGSAWAKIYGVNVDGKNHFTYSIKPELVGQOQIIPWNFFLLMNSWKITGP 202

QY 182 ALVTGNTVILKTAESTPLSALYVSKYIPQAGIPPPVINTVSGFISVVEATINHHKIKKV 241

Db 203 ALATGNTVVIKPAFTPLSALFASQVQAGNIPAVVNLPSGPPVGEPISAHCEVKKI 262

QY 242 APTGSTATGRHHYQSAAGLKKVYIDLGQKSNIVHFAELKAVQVNIILGSIYNSGWC 301

Db 263 APTGSTATGRHHMKVAADPVKKVTELOGKSNIVFADADLDKAVKNIAFGIEYNSGVC 322

QY 302 CAGSRVTVESIVDKFLEEFKAASESIKVGDPDESTFOGAOTSOMLNKILKYVDIGKN 361

Db 323 CAGSRVTVQTVVFFVFIKRYTESLPVNDPDEEVFOGATSKQKLRKILDYDVAKS 382

QY 362 EGATLITGGRIAGSKGYFIKPTVFGDVQKEDMRIVKKEIFGPVTVTKFKSADEVINMAND 421

Db 383 EGATLITGGRIAGSKGYFIKPTVFGDVQKEDMRIVKKEIFGPVTVTKFKSADEVINMAND 442
 QY 422 SEYGLAAGIHTSNITALKVADRVNAGTVMINTYNDPHFHAVPFGGPNASGLGRENMSVDAL 481
 Db 443 SEYGLAAGIHTSNITALKVADRVNAGTVMINTYNDPHFHAVPFGGPNASGLGRENMSVDAL 502
 QY 482 QNYLOQKAVRAKLD 495
 Db 503 SNYTDTKSVRIAD 516

RESULT 3

S60929

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.4) YPL061w yeast (Saccharomyces cerevisiae)

N:Alternate names: protein IPE9w

C:Species: Saccharomyces cerevisiae

C>Date: 15-Feb-1996 #sequence_revision 31-Jan-1997 #text_change 04-Jun-2002

C:Accession: S60929

R:Winkler, F. Almond, A. Hossley, B. Porfiri, N. Friesen, J. D. Hall, J. Storms, R.

submitted to the EMBL Data Library, October 1995

A:Reference number: S60921

A:Accession: S60929

A:Molecule type: DNA

A:Residues: 1-500 <WIN>

A:Cross references: EMBL:U08206; NID:G60722; PDB:1AAR64612; PDB:1U09001; GSPDB:G00000

C:Genetics:

A:Gene: SGD:ALD6; MIPS:YPL061w

A:Cross references: SGD:S0005982; MIPS:YPL061w

A:Map position: 16L

C:Function:

A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water

A>Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD, oxidoreductase

F:64-326/Domain: aldehyde dehydrogenase homology ALD6

F:272,306/Active site: Glu, Cys #status predicted

Query Match 56.9%; Score 1452; DB 1; Length 500;

Best Local Similarity 54.8%; Pred. No. 9.8e-88;

Matches 272; Conservative 92; Mismatches 128; Indels 4; Gaps 4;

QY 4 HLPMTVPILK LPLNGLEYEPTGLFINNKFVPSKQNTFEVINPSTFEFTCHYEGRED 60

Db 5 HPTAEFVKTITLNGLEYEPTGLFINNKFVPSKQNTFEVINPSTFEFTCHYEGRED 64

QY 61 VEFVAAAKAFESNLSWNIIDLGQKALYRLAFILPQKQDVIASIETLQNGKATSSSG 120

Db 65 VEFVAAKAFESNLSWNIIDLGQKALYRLAFILPQKQDVIASIETLQNGKATSSSG 124

QY 121 DVDLVINYLKSSAGFAUKIDGRMIDTGRTHSYTKRQPGVGGQIIPWNFFLLMNAWKIA 180

Db 125 DVTIATNCLRDAAAYAKVNGRTINTGQYMNFTTLEIGVGGQIIPWNFFLLMNAWKIA 184

QY 181 PALVTGNTVILKTAESTPLSALYVSKYIPQAGIPPPVINTVSGFISVVEATINHHKIKKV 240

Db 185 PALAMNVGILKPAFTPLSALFASQVQAGNIPAVVNLPSGPPVGEPISAHCEVKKI 244

QY 241 VAPTGSTATGRHHYQSAAGLKKVYIDLGQKSNIVHFAELKAVQVNIILGSIYNSGWC 299

Db 245 LAFTGSTEIVKSKAVAVSSSNLKKIIELEGKSNIVFADADLDKAVKNIAFGIEYNSGVC 304

QY 300 VVAGSRVTVESIVDKFLEEFKAASESIKVGDPDESTFOGAOTSOMLNKILKYVDIGKN 358

Db 305 PNSSGSPVYVQGGIYHETDAAFKATVETIKVNSGPPVGEPISAHCEVKKI 364

QY 359 GNEZATLITGGRIAGSKGYFIKPTVFGDVQKEDMRIVKKEIFGPVTVTKFKSADEVIN 418

Db 365 GKKEGAKILGQKQVADYFIRPVVYVNLDMKIVKELDGVVIVAKRILLEGVEM 424

QY 419 ARISYGLAAGIHTSNITALKVADRVNAGTVMINTYNDPHFHAVPFGGPNASGLGRENMSVDAL 478

Db 425 ANSEFGLSGLEFESLSEFVAPMLFAGVWNIYNIHESVVEGVKSGYEGEMGE 484

QY 479 DALONYLOVKAIRAKL 494
Db 485 EYHAYTEVKAVIRKL 500

RESULT 4
A29055
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - *Emmericella nidulans*
C:Species: *Emmericella nidulans*, *Aspergillus nidulans*
C:Date: 30-Jun-1989 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: A29055
R.Pickett, M.; Gwynne, D.I.; Buxton, F.P.; Elliott, R.; Davies, R.W.; Lockington, R.A.;
Gene 51, 217-226, 1987
A:Title: Cloning and characterization of the aldA gene of *Aspergillus nidulans*.
A:Reference number: A29055; MIM:87248080; PMID:303652
A:Accession: A29055
A:Molecule type: DNA
A:Residues: 1-497 (PIR)
A:Cross-references: GR-M16197; NID-g168010; PIDN:AAA33293.1; PID:g168011
A:Experimental source: strain Glasgow F98c4
C:Genetics:
A:Gene: aldA
A:Map position: VIII
A:Introns: 21/2; 422/1
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; NAD; oxidoreductase
F:57-318/Domain: aldehyde dehydrogenase homology <ALD>
F:264,298/Active site: Glu, Cys #status predicted

Query Match 55.4%; Score 1412.5; DB 1; Length 497;
Best Local Similarity 54.5%; Pred. No. 3.8e-85;
Matches 270; Conservative 88; Mismatches 132; Indels 5; Gaps 3;

QY 1 MSLLPMTVPKLPNGLEYEQFTGLFINNKEVPSKQNKTEFEVINFSTEEIEICHIEGKEDD 60
Db 1 MSLLPMTVPKLPNGLEYEQFTGLFINNKEVPSKQNKTEFEVINFSTEEIEICHIEGKEDD 60

QY 61 VELAVQADAPAFNSNWNIDIPDRKALYPIAFILPQKIVKIVIASIEILONGKAISSSRG 120
Db 58 VDVAVAAARAF-EGEVRQVTPSPKGLINKLADLMERDIDTAAIESUNGKAFPMKV 116

QY 121 DVDLVINYLKSSAGFADKIDGRMTDGRTHFSYTKRQPLGVCGGIIIPWNPPLLMWAKIA 180
Db 117 DLANSIGCLRYAGWADKTHGQTDITNPETLTITRHPVGVCGGIIIPWNPPLLMWSWKIG 176

QY 181 PALVTGNTVWLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIVVEATINHPKIRK 240
Db 177 PAVAAAGNVVLKTAQTPLSALYAAKLIKEAPFAGVINIVSGFGRTAGAAISSHMDIK 236

QY 241 VAFTGSTATGHIYQSA-AGLKVVTELGKSNIVFADAELKKAQVNLILGIYNSGE 299
Db 237 VAFTGSTLVGTLQAAKSNLKVVTELGKSNIVFDDADLUNALSWANFQIFPNHGQ 296

QY 300 VCCAGSRVYVESYDKPIEFIEFKAASIESIKVGOPFDGSTFQGAQTSQMQLNKILKYVDIG 359
Db 297 CCCAGSRILVQEGYDKVAKFKEKAKQNVGNPFEDTQGGVQVSOLOFDRIMEYING 356

QY 360 KNEGATLTGGERIGSGKGYFTIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSKADEVINMA 419
Db 357 KKAGATVATGDKHNGEYFQIPVFIQVTDVTSUMKIAQEEIFGVVTVTKFSKADEVINMA 416

QY 420 NDSYGLAAGLHTSNITALKVADRVNAGTVWINTYDFHHAHPFGGFGNAGSLGKREMSVD 479
Db 417 NSTDYGAAAVHTKVNITATIRVSNALKAGTVWNNYISYQAPFGGFGKSGLGLRELSY 476

QY 480 ALQNYLOVKAIRAKL 494
Db 477 ALENTYQIKTVHYRL 491

RESULT 5

S43114

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - fungus (*Cladosporium herbarum*)C:Species: *Cladosporium herbarum*

C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 03-Jun-2002

C:Accession: S43114

P.Abbat, G.; Oberkofler, H.; Simon, B.; Lehenauer, E.; Unger, A.; Kandler, D.; Pril

submitted to the EMBL Data Library, March 1994

A:Description: Molecular characterization of allergens of *Cladosporium herbarum* and A

A:Reference number: S43108

A:Accession: S43114

A:Molecule type: mRNA

A:Residues: 1-495 (ACH)

A:Cross-references: EMBL: X74228; NID: g457623; PIR: AAF072.1; PIR: g457625

A:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:56-317/Domain: aldehyde dehydrogenase homology ALD

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water
 A:Pathway: ethanol catabolism
 A>Note: enzymes with this activity are involved in diverse metabolic pathways in various C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase
 F:1-19/Domain: transit peptide (mitochondrion) #status predicted <NIP>
 F:23-51/Product: aldehyde dehydrogenase (NAD⁺) 2 #status predicted <MAL>
 F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>
 F:287,321/Active site: Glu, Cys #status predicted
 F:474/Binding site: NAD (Cys) #status predicted

Query Match 50.0% Score 1275.5 DB 1 Length 519;
 Best Local Similarity 51.0% Pred No. 4e-76;
 Matches 253; Conservative 84; Mismatches 146; Indels 13; Gaps 6,

QY 7 TVPIKLPNGLEVPQ-----TGLFINNKVPKSKQNKTEVINPSTEEIECHLYCKEDVE 62
 DB 24 TSAVAPN-----HQPEVFCNQIFINNEWHDAVSRKKTFTVNPSTGEVICQVAGCKEDVD 79
 QY 63 EAQQAADRAFSNCS--WNQIDIDGKALYKLAFLLEKQKVIASTETLQNGKA-ISSSKG 120
 DB 80 KAVKARAFAQLGSPWRMDASDRGLLYPLADLIERDPTVYLALETLDNGKPYVISYLV 139
 QY 121 DVLVINYLKSSAGFADKIDGRMIDTSTHESYTKPQLGVQGGIIPWNEPFLMWAKKIA 180
 DB 140 DLDMLKCLPYVYAGWADKYHCKTIPIDGDFESSYTHRPVGVGCGIIPWNEPFLMWAKKLG 199
 QY 181 PALVTGNTVLTAESTPLSALYSKYIPIQAGIPPGVINIVSGPKIIVVETINHPKIK 240
 DB 200 PALATGNVVMKVAEQTEPLTALYVANLKEAGFPFGVNVIVPGPGTAGAIAASHEDVDK 259
 QY 241 VAFTRSTATGRHIYGSAA--AALKKVFLIDGKSFNIFVAFBAELKKAVQNIILGIYVNSG 298
 DB 260 VAFTRSTEVG--HLIVAAAGSSNLKVVLELDGKSNILMSIAQMWAVEVAHAFALPNQG 318
 QY 299 EYVQAGSRVYVEESLYDKTEIEEFKRAESIKVSGDFEETFGQAGISQMIKILKYVDI 358
 DB 319 QQQCAATSEFVGENYDEFEVERSVAFKSRVGVNPNFSPTRQGPQVDEIQPKKILGYIKS 378
 QY 359 GKNEATLITGCEPLGSKGYFTKPVFVGVKLEHPIKKEELFGPVVTVTKHKSADRVINM 418
 DB 379 QQERAKLGGGTAARAGRYFIPQIVFVGVKPMETIAKEFIQGVGMQILAFKTEEVVSP 438
 QY 419 ANDSEYGLAAGIHISNITALKVADKVNAGTWINTYNDGFHVAFFPGSGFNASSGLSGMSV 478
 DB 439 ANDSKYGLAAAVFTKDLKANYLSQALQAGIWIWNCYDVFGAUSPPGSGYKMSGSPRELGE 498
 QY 479 DALQNYLQVKAVRAKL 494
 DB 499 YGLQAYTEKVTVTVKV 514

RESULT 9

S03564

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 28-Feb-1990 #sequence_revision 31 Jan-1997 #text_change 03-Jun-2002

C:Accession: S03564, A27713; S12903; S17432

R:Farres, J.; Guan, K.L.; Weiner, H.

Eur J Biochem 180, 67-74, 1989

A:Title: Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenase

A:Reference number: S03564; MIMR:R4210865; PMID:2450003

A:Accession: S03564

A:Molecule type: mRNA

A:Residues: 1-519 <FAP>

A:Cross-references: EMBL: X14977, NID: g55604; PDB: CAA33101 1; PDB: g55605

A:Experimental source: strain Sprague Dawley; liver

R:Farres, J.; Guan, K.L.; Weiner, H.

Biochem. Biophys. Res. Commun. 150, 1083-1087, 1988

A:Title: Sequence of the signal peptide for rat liver mitochondrial aldehyde dehydrogenase

A:Reference number: A27713; MIMR:R8134217; PMID:3342060

A:Accession: A27713

A:Molecule type: mRNA

A:Residues: 1-519 <FAL>

A:Cross-references: GB: M39936, NID: g302847, PDB: AAA071913, PDB: g302848

R:Guan, K.L.; Pallwal, R.; Kaftan, E.; Bawa, R.

FEBS Lett. 273, 215-218, 1990

A:Title: A mitochondrial protein fraction catalyzing transport of the K(+) analog Tl(+)

A:Reference number: S12903; MIMR:R421084; PMID:1649808

A:Accession: S12903

A:Molecule type: protein

A:Residues: 323-430 <EDW>

R:Geng, J.; Weiner, H.

Arch. Biochem. Biophys. 289, 214-222, 1991

A:Title: Purification and characterization of catalytically active precursor of rat l

A:Reference number: S17432; MIMR:R421084; PMID:1649808

A:Accession: S17432

A:Molecule type: protein

A:Residues: 1-19 <FEN>

C:Genetics:

A:Genome: nuclear

C:Complex: homotetramer

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A>Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta

F:1-19/Domain: transit peptide (mitochondrion) #status predicted <NIP>

F:20-519/Product: aldehyde dehydrogenase (NAD⁺) 2 #status predicted <MAL>

F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>

F:287,321/Active site: Glu, Cys #status predicted

F:474/Binding site: NAD (Cys) #status predicted

Query Match 50.0% Score 1274.5 DB 1 Length 519;

Best Local Similarity 50.8% Pred No. 4.7e-76;

Matches 252; Conservative 85; Mismatches 146; Indels 13; Gaps 6;

QY 7 TVPIKLPNGLEVPQ-----TGLFINNKVPKSKQNKTEVINPSTEEIECHLYCKEDVE 62

DB 24 TSAVAPN-----HQPEVFCNQIFINNEWHDAVSRKKTFTVNPSTGEVICQVAGCKEDVD 79

QY 63 EAVQAADRAFSNCS--WNQIDIDGKALYKLAFLLEKQKVIASTETLQNGKA-ISSSKG 120

DB 80 KAVKARAFAQLGSPWRMDASDRGLLYPLADLIERDPTVYLALETLDNGKPYVISYLV 139

QY 121 DVLVINYLKSSAGFADKIDGRMIDTSTHESYTKPQLGVQGGIIPWNEPFLMWAKKIA 180

DB 140 DLDMLKCLPYVYAGWADKYHCKTIPIDGDFESSYTHRPVGVGCGIIPWNEPFLMWAKKLG 199

QY 181 PALVTGNTVLTAESTPLSALYSKYIPIQAGIPPGVINIVSGPKIIVVETINHPKIK 240

DB 200 PALATGNVVMKVAEQTEPLTALYVANLKEAGFPFGVNVIVPGPGTAGAIAASHEDVDK 259

QY 241 VAFTRSTATGRHIYGSAA--AALKKVFLIDGKSFNIFVAFBAELKKAVQNIILGIYVNSG 298

DB 260 VAFTRSTEVG--HLIVAAAGSSNLKVVLELDGKSNILMSIAQMWAVEVAHAFALPNQG 318

QY 299 EYVQAGSRVYVEESLYDKTEIEEFKRAESIKVSGDFEETFGQAGISQMIKILKYVDI 358

DB 319 QQQCAATSEFVGENYDEFEVERSVAFKSRVGVNPNFSPTRQGPQVDEIQPKKILGYIKS 378

QY 359 GKNEATLITGCEPLGSKGYFTKPVFVGVKLEHPIKKEELFGPVVTVTKHKSADRVINM 418

DB 379 QQERAKLGGGTAARAGRYFIPQIVFVGVKPMETIAKEFIQGVGMQILAFKTEEVVSP 438

QY 419 ANDSEYGLAAGIHISNITALKVADKVNAGTWINTYNDGFHVAFFPGSGFNASSGLSGMSV 478

DB 439 ANDSKYGLAAAVFTKDLKANYLSQALQAGIWIWNCYDVFGAUSPPGSGYKMSGSPRELGE 498

QY 479 DALQNYLQVKAVRAKL 494

DB 499 YGLQAYTEKVTVTVKV 514

RESULT 10

DEH032
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial [validated] - human
 N:Alternate names: aldehyde dehydrogenase f2; aldehyde dehydrogenase I
 C:Species: Homo sapiens (man)
 C>Date: 31-Mar-1988 #sequence, revision 17-Nov-1995 #text, change 03-Jun-2002
 C:Accession: A29975; S00804; A23503, A27509, A26743, I39432, I39433, I39430
 R:Hsu, L.C.; Benda, P.E.; Yoshida, A.
 Genomics 2, 57-65, 1988
 A:Title: Genomic structure of the human mitochondrial aldehyde dehydrogenase gene.
 A:Reference number: A29975; MUID:88256152; PMID:2838413
 A:Accession: A29975
 A:Molecule type: DNA
 A:Residues: 1-517 <HSU1>
 A:Cross-references: GB:M26760; NID:q178397; FIDN:AAA51694.1; PID:q178398
 R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
 FEBS Lett. 233, 440, 1988
 A:Reference number: S00804
 A:Accession: S00804
 A:Molecule type: mRNA
 A:Residues: 1-6, 'APA', 10, 'P', 13-49, 'UPA'
 A:Cross-references: EMBL:X05409; NID:928605
 A:Note: Contribution to A26743
 R:Hempel, J.; Kaiser, R.; Joernvall, H.
 Eur. J. Biochem. 153, 13-28, 1985
 A:Title: Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, difference
 A:Reference number: A23503; MUID:86055846; PMID:4065146
 A:Accession: A23503
 A:Molecule type: protein
 A:Residues: 'A', 19-517 <HEM>
 A:Note: The sequence shown is presumably that of the mature protein, however, the number
 R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
 Nucleic Acids Res. 15, 3179, 1987
 A:Title: Isolation and sequence analysis of a full length cDNA clone coding for human mi
 A:Reference number: A27509; MUID:87174836; PMID:3562250
 A:Accession: A27509
 A:Molecule type: mRNA
 A:Residues: 1-6, 'AWPAA', 10, 'P', 12, 'VS', 15, 'RHPR', 21, 27-79, 'REGPRG', 86-336, 'V', 338-517
 A:Cross-references: EMBL:Y00109; NID:928607; FIDN:CAA6290.1; PID:q28608
 R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
 FEBS Lett. 215, 233-236, 1987
 A:Title: Evidence for a signal peptide at the amino-terminal end of human mitochondrial
 A:Reference number: A26743; MUID:87219091; PMID:3582651
 A:Accession: A26743
 A:Molecule type: mRNA
 A:Residues: 1-6, 'AWPAA', 10, 'P', 12, 'VS', 15, 'RHPR', 21, 27-79, 'REGPRG', 86-336, 'V', 338-517
 A:Cross-references: EMBL:X05409; NID:928605
 A:Note: This sequence is revised in reference S00804
 R:Hsu, L.C.; Tani, K.; Fujiyoshi, T.; Kurachi, K.; Yoshida, A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3771-3775, 1985
 A:Title: Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.
 A:Reference number: I39432; MUID:85216574; PMID:2987944
 A:Accession: I39432
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 119-336, 'V', 338-517 <HSU2>
 A:Cross-references: GR:K0101; NID:q178395; FIDN:AAH59500.1; PID:q178396
 A:Note: thirty-three tryptic peptides were also sequenced
 R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
 Alcohol 2, 103-106, 1985
 A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
 A:Reference number: I39431; MUID:85252089; PMID:4015823
 A:Accession: I39431
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 119-336, 'V', 338-517 <YOS>
 A:Cross-references: GB:M26760; NID:q178397; FIDN:AAA51694.1; PID:q178398
 R:Agarwal, D.P.; Goedde, H.W.
 Isozymes Curr. Top. Biol. Med. Res. 16, 21-48, 1987
 A:Title: Human aldehyde dehydrogenase isozymes and alcohol sensitivity.
 A:Reference number: I39430; MUID:87779033; PMID:4610692
 A:Accession: I39430
 A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA
 A:Residues: 214-215, 'S', 217, 'K', 219-246, 'P', 248-436, 'V', 438-426, 'EVGILRGRGWNHSGQFHVRAG
 A:Cross-references: GB:M54931; NID:q178391; FIDN:AAA2825.1; PID:q178392
 A:Note: the sequence is misidentified as aldehyde dehydrogenase 1
 C:Genetics:
 A:Gene: GDB:ADH2
 A:Cross-references: GDB:112608, OMIM:136650
 A:Map position: 12q24.2:12q24.2
 A:Insertions: 874, 774, 1267, 1472, 180, 5, 227, 5, 267, 7, 300, 9, 612, 43, 6, 40, 22
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
 C:Superfamily: aldehyde dehydrogenase (NADP+), aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta
 F:1-17/Domain: transit peptide (mitochondrion) #status predicted -SP
 F:18-517/Product: aldehyde dehydrogenase (NADP+) #status experimental -MA
 F:75-339/Domain: aldehyde dehydrogenase homology -ALDH
 F:211-291/Domain: NAD binding #status predicted -NAD
 F:285/Active site: #status predicted
 F:319/Active site: Cys #status experimental
 F:472/Binding site: NAD (Cys) #status predicted

Query Match 49.2% Score 1254.9; DB 1; Length 517;
 Best Local Similarity 50.4%; Pred. No. 1 to 74;
 Matches 249; Conservative 86; Mismatches 149; Indels 11; Gaps 5;
 QY 7 TVPIKLPNGLEYEQP-----TGLPIINNKVPISKONKPEVININSTEETICHYERDEIVE 62
 LB 1
 LB 22 TQAVEAFN - QQTEVECNQIFINNEWDVAVSRKTTTTFVPSIGEVITQVAERKEIVD 77
 QY 63 EAVQAAADRAFPNCS-WNGIDPDRKALYLAELPQDKVIASLETINIKKA-LSSSRG 120
 LB 1
 LB 78 KAVKAAKAAAFGLSSPPKPKMLASHKSHLLNKLALTEKPEETLAALETINIKKPPVSYLV 147
 QY 121 DVDLVINYLKSSAGFAURKIDGRMIDGRTHFSYTRKQPLAVGQITPNNFELIMWAKIA 180
 LB 1
 LB 138 ELMQVLEKELKYAGAWAKLQKGLKFLQGLDFAVLRKELVWVGGQITWNLTEMGAWKAG 197
 QY 181 PALVGNIVVKEIAESPLSALYVSKYDPAVLPPEVININSTEETICHYERDEIVE 240
 LB 1
 LB 198 PALATENVVMKVAIVTPIALYVANG-TEAVTPPEVNVN;VNSPTPA-MAALSHPIVHP 257
 QY 241 VAFSTATGSEPTFYCA-AAAGFPVVEETGKPEHIVIAALPKAVN;ETITLYNYPPE 299
 LB 1
 LB 258 VAFSTGTEICRVQVAAAGSSNLKRVLELQKSNINLSDAIMQWAVFQAHPFENQGG 317
 QY 355 VQVAFSPVVEESTYVFFPEPKAASNLKYVQDPPESIFGQALISLQMLNFKHLYVDIG 359
 LB 1
 LB 338 QVQASSKPEVQEDTDLAVLRGVAFKSKVGNINTEKELQSGQVHRLQKELDYTHG 377
 QY 360 KNHGATLTGGERLQSKGYFKPTVTPVAVKEMIRVKEETFGQVVFVKPSALEVINMA 419
 LB 1
 LB 378 KQEGAKLGGGGLAIBRGYFLOPTVEGVQVQVMTLAKETITGQWMLKPKTEDEVVIRA 447
 QY 420 NUSEYGLAAGIHISININIAKAVADVKNACTVNTINNIHRAVPEPEFNASAKREMSVD 479
 LB 1
 LB 448 NNSTYGLAAAVFTKDLKANYLSQALQACTVWNVYDFVFCAGSDPPGQYKMSGSEKLEY 497
 QY 480 ALQNYLOVKAVRAKL 494
 LB 1
 LB 498 GLQAYTEVKTVTKV 512

RESULT 11

B69614
 aldehyde dehydrogenase dbas - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence, revision 05-Dec-1997 #text, change 20-Jun-2000
 C:Accession: B69614
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bor
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Cardenas, V.; Carter, N.M.;

A: Ehrlich, S. D., Frensch, P. L., Entian, K. D.: *Biotechnology*, 1: Fabrot, C.; Ferrari, E. Nature 390, 249-256, 1997

A: Authors: Fiedler, G.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C. R.; Heide, A.; Hiller, H.; Holsappel, S.; Hosono, S.; Hui, M. F.; Keckler, P.; Koningstein, G.; Krogl, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Maue, Y. M.; Ogawa, K.; Ohtawa, A.; Oudry, R.; Park, S. H.; Paro, V.; Pohl, T. M.; Portet, R.; Rieger, M.; Rivollat, C.; Roch, E.; Roche, R.; Rose, M.; Sadaler, Y.; Sato, T.; Stanton, A.; Thiers, S.; Schuetz, F.; Stoffe, F.; Sukiguchi, J.; Sekowska, A.; Serron, A.; Winters, M.; Tanaka, A.; Tanaka, T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiyama, K.; Winters, M.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.; Zumbach, E.; Yoshikawa, H.; Zumbach, E.; Yoshikawa, H.

A: Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A: Reference number: A63580; MIM: 6844603; PMID: 9484377

A: Accession: H69614

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-495 <KUN>

A: Cross-references: 5B:290114; GR:AL009126; NID:9234230; PION: CAB13423 1; PID: 9263424

A: Experimental source: strain 168

C: Genetics:

C: Gene: dhas

C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C: 58-329/Domain: aldehyde dehydrogenase homology: ALD

Query Match: 49.1%, Score 1252.5, DB 2, Length 495,
Best Local Similarity 49.0%, Prod No 12074,
Matches 243, Conservative 92, Mismatches 158, Indels 3, Gaps 2,

QY 1 MSHLPMTVPKLPNCLEVEOPTGLFNNKVPSPKQNTFVNPSTPEELCHYEGRED 60
DB 1 MSSLMQVTKETFLQGTKK--LYIDGKEVPASGASGETPDPATGETLMTLYEAQAD 58
QY 61 VEEAVQADRAFSGNSWNGDIPDRGKALYRLAFLPQKQDVIASTFLDNGKAS--SSR 119
DB 59 VDKAVKAPKAFDQGEWRTSPASRSLMYKLADLMEHKTLELAQLETLNDSKPINETTN 118
QY 120 GDVLIVNLYKSSAGFADKIDGRMIDTGRTHFSYTRQPLGVCQGIIPNFPLLMAWKI 179
DB 119 GDPLAIEHMYVAGWCTKITGQTPVSGAVFNTVTRHVPVGVVQGIIPNFPLLMAWKI 178
QY 180 APALVTGNTVTLTAESTPLSALVSKYIPQAGIPGVINVSFGKIVVEATINIPKIK 239
DB 179 GAALATGCIIVLKABQTPSALYLAELIDQAFPAVINLIPGDEAGEALTNHEAVD 238
QY 240 KVAFGTGSTATGRHIVQSAAGLKVTLLELCKSPNIVFADAELKAVQNIILGIYNSGE 299
DB 239 KIAPFTGPEICKKIMSTAAKSIKRVTLLELCKSPNILLDANLKAIFGALNGVMFNQGO 298
QY 300 VCCAGSRVYVEESTYDKFIEFKAASESIKVGDPFDETFQGAQTSOMQLNKLKVDIG 359
DB 299 VCCAGSRVYVTHQVYHAKVWPMASVAFSLWQAGLHKTGICPLVSKQEHVSLYIQKG 358
QY 360 KNEATLITGGERLGSQGYFTKPTVFGVDEKEDMPVYKRETFGVPVTVTRFKSADEVINNA 419
DB 359 KDEGAKAVTGSCGFPEAGYFVAPVTFANVEDMTIAKEEFGPVLTAIYEIVDEVIERA 418
QY 420 NDESEYGLAAGIHTSNITALKVADKVNAGIWINIYNDPHHVPFGGFGNAGLGRMSVD 479
DB 419 NHSEYGLAAGLWENVKQAHYIADRLQAGTWWVNCYNVFDDAASPFGRYKQSLGRMGSY 478
QY 480 ALONYLOVKAVRAKLD 495
DB 479 ALDNTTEVSKSVNLE 494

RESULT 12

A40872

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 5 precursor, mitochondrial - human

C: Species: Homo sapiens (man)

C: Date: 27-Mar-1992 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002

C: Accession: A40872

R: Hsu, L.-C.; Chang, W.-C.

J: Biol. Chem. 270, 1257-1265, 1995

A: Title: Cloning and characterization of a new functional human aldehyde dehydrogenase

A: Reference number: A40872; MIM: 6126241; PMID: 2061311

A: Accession: A40872

A: Molecule type: DNA; mRNA

A: Residues: 1-517 <HSD>

A: Cross-references: GR:653467; NID: 9337184; PION: AAA96830 1; PID: 91264008

A: Note: 85-Ala and 107-Leu were also found

C: Comment: based on similarity, this form is assumed to be mitochondrial and to have

C: Genetics:

A: Gene: GDB:ALDH5; ALDHX

A: Cross-references: GDB:128788

A: Map position: 9p13-9p13

A: Note: the only initiation occurs before the initiator codon

C: Complex: homotetramer (probably)

C: Function:

A: Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water

A: Pathway: ethanol catabolism

A: Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C: Keywords: alcohol metabolism, homotetramer, liver, mitochondrion, NAD, oxidoreducta

C: 1-20/Domain: transit peptide (mitochondrion) #status predicted <NTP>

C: 21-517/Product: aldehyde dehydrogenase (NAD+) 5 #status predicted <NAT>

C: 75-333/Domain: aldehyde dehydrogenase homology: ALDH5

C: 285,319/Active site: Glu, Cys #status predicted

Query Match: 49.1%, Score 1252.5, DB 1, Length 517,
Best Local Similarity 50.6%, Prod No 13074,
Matches 248, Conservative 85, Mismatches 148, Indels 9, Gaps 6;

QY 9 PRLPNSIPEYQPHPLPNNKVPFSKJNTFFVINISTFEELPHVFGPEIDVFFAVQAA 49
DB 28 PRLNPDVLPXNQ--LLEINNEWQAVSKKIFPVPNPLGEVIGHVAGGKALVHAKVAA 83
QY 69 LPKASNGS--WNGIHPHLPKALYPLAFLPQDKIVIASLFTLQNKATSSSP3-DVDIVL 126
DB 84 PEVPLGSPPPMPASFPGLPLNPLADLVEPDPVYLSLETLDNGKPFQESTALDDEV 143
QY 127 NULKSSAGFALKIDSPMIDTSPHPSYIKKQPEVQVQGIIPNFPNPLMMWAKIAPALVTG 186
DB 144 KVVYFAGWAKWKGKTIPTMGQHPGCTRHEPVGVQVQGIIPNFPNPLVMQWKLAPALATG 203
QY 187 NTVLKTAESTPLSALVSKYIPQAGIPGVINVSFGKIVVEATINHPKIKKVAFTGS 246
DB 204 NTVMKVAEOTPLSALYLAELIDQAFPAVINLIPGDEAGEALTNHEAVD 263
QY 247 TATGRHIVQSA--AGLKVTLLELCKSPNIVFADAELKAVQNIILGIYVNSGEVCCAG 304
DB 264 TEVG-HLIQKAGDSNLRVTLLELCKSPNIVFADAELKAVQNIILGIYVNSGEVCCAG 322
QY 305 SPVYVESIYDFIEEFKAASESIKVGDPFDETFQGAQTSOMQLNKLKVDIG 364
DB 323 SPTEVPFSTVNFPTVEKAKQPKVGNPPELDTQGGPQVVKFQGFVYIQLVQKGA 382
QY 365 TLITGGELPLGSKGYFETPVEGVFEMPIVKEEFGVPVTVTRFKSADEVINWANDISEY 424
DB 383 KLICGGEPFSGPFFIKPTVFGVQVNDMPITAKETFPQVDPLEPKFKKIEEVEERANTY 442
QY 425 GIAAGITSNITALKVADKVNAGIWINIYNDPHHVPFGGFGNAGLGRMSVD 484
DB 443 GLAAAVPIKLDKAMVFTQALQAGIVWVNIYVITGHPGSGKSGKELGKELKAY 502
QY 485 LQVKAAPAKL 494
DB 503 TEVTVTVTKV 512

RESULT 13

S09030

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial - bovine

C: Species: Bos primigenius taurus (cattle)

C: Date: 30-Jun-1992 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002

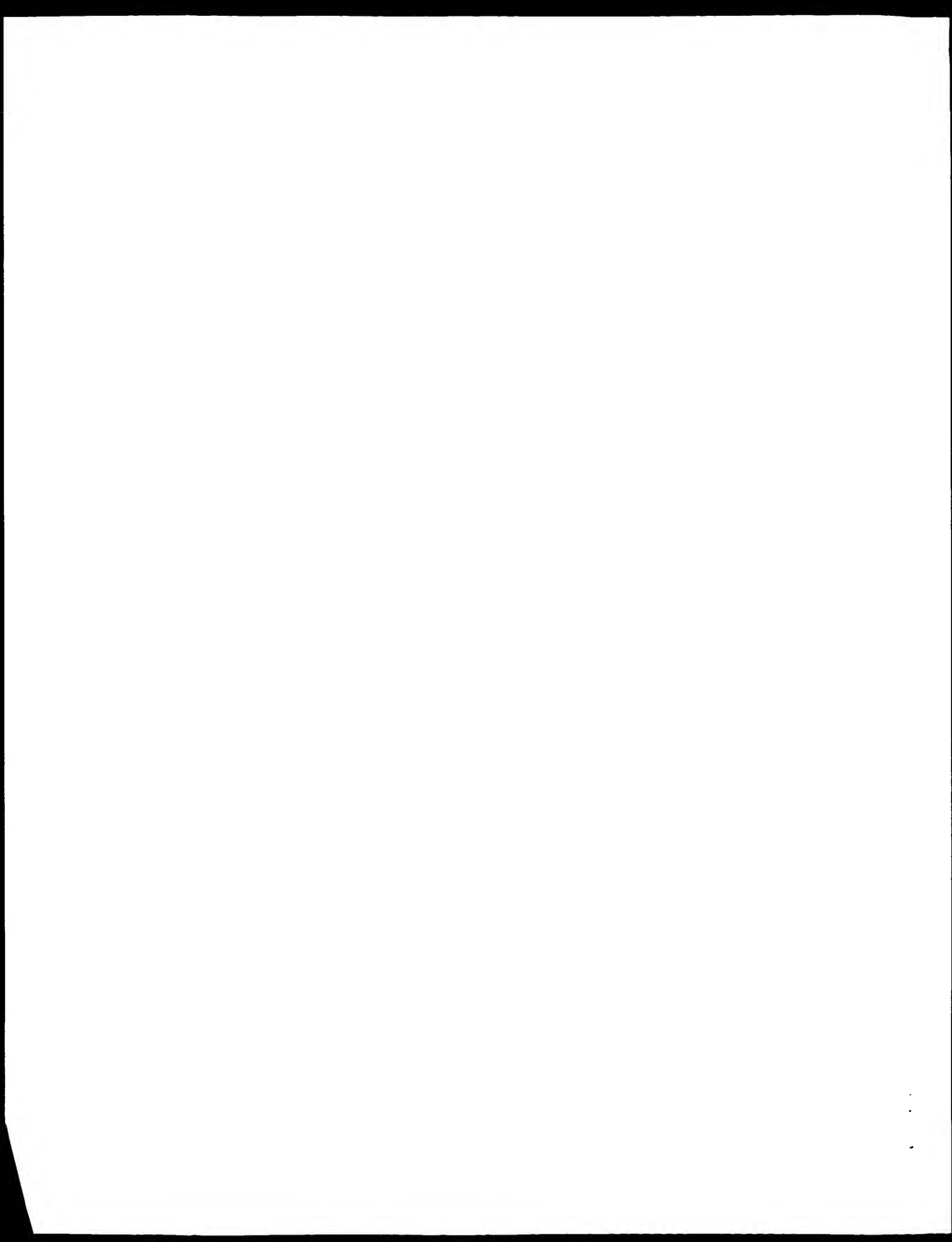
C: Accession: S09030; S03565; P00543

N: Alternate names: aldehyde dehydrogenase homolog
 C: Species: Zea mays (maize)
 C: Entry: AF039834; accession: AF039834; date: 11-May-2000
 C: Accession: T03983
 R: Cui, X.; Wise, R.P.; Schnable, P.S.
 Science 272, 1334-1336, 1996
 A: Title: The r12 nuclear restorer gene of male-sterile T-cytoplasm maize.
 A: Reference number: 215177; MUID: 96243131; PMID: 8650543
 A: Accession: T03983
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-549 <CUI>
 A: Cross-references: EMBL: U03983; NID: 96243131; FID: 96243131
 A: Experimental source: strain B73
 C: Genetics:
 A: Map position: 9
 A: Note: r12
 C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 F: 107-370/domain: aldehyde dehydrogenase homology <ALD>

Query Match 48 38; Score 1232.5; DB 2; Length 549;
 Best Local Similarity 48.1%; Pred. No. 2.9e-73;
 Matches 244; Conservative 51; Mismatches 143; Indels 23; Gaps 5;

QY	7	TVPIK-----LPNGLE-----YEOP-----TGLFINNKFPVPSKQNTFEVINPS	45
DB	33	TVPADGMIRLLPGVLPFSTAAVFPETPSVHVNYTKLLINGNFVDSASGKTFTLDP	92
QY	46	TEETICHLYEKREDVVEAVQAAKAFNSGWSNGIDPDKKALYKLAELIEQDKDVAS	105
DB	93	1GEVIAHVAGDAEDINRAVAAAKAFDESPWFKMTAYEKRILLKFAADLIEKHNDLAA	152
QY	106	IELDNGKAT-SSSRGQDVLVINYLSKSGAFADKIDGPMIDGPTHFSYTKRQPIGVCGQ	164
DB	153	LETWNGKPFQAAAEVPMVAFIMBYAATWAKTHSIVPAFGPHHVQILHEPISVAGQ	212
QY	165	IIPWNPFLLMWANKIAPALVTGNTVLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGF	224
DB	213	IIPWNPFLMYAWKVPALACGNTLVLKTAEQPLSALYISKLLHEAGLPEGVWVWSGF	272
QY	225	GKIVVEATNHPKIKKVAFTGSTATGPHIYQSA-AGLKVKVILELGGKSPNIVFADAEK	283
DB	273	GPTAGAAASHMDVKIAFTGSTDGKIILELAASKNLKTVILELGGKSPFIIMDDADV	332
QY	284	KAVENITIGIYNSGEVTCAGSRVVEESTYKFEIEEFKAASESKVGGPFDESTFQGAQ	343
DB	333	HAVELAHFALFFNGUNQACASPTFVHEKVFDEFKAKAPALKVVGTPPEKGVESHPQ	392
QY	344	TSQMLNKLKLYVDIGNEGATLITGGERIGSGYFIKPTVFGDVKEDMRIYKEEIPGPV	403
DB	393	IDDEQFNKILLPYIPYGVGGGATLVTSRQFQSHKGFYIQPTIFSDVGLGKMKIAGEEIPGPV	452
QY	404	VTVTKEKSADEVINWANDSEYGLAAGIHTSNTALKVADRVNAGTVWINTYNDEHHAVP	463
DB	453	QSILKFKDLNEVIRKASQYGLAAGVFTNSLDTANTLTAKRAGTVWVNCDFVDAAIP	512
QY	464	FGGNASGLGSEMSVDALQNYLQVKAV	490
DB	513	FGGYKMSGIGREKGVDSILKNYLQVKAV	539

Search completed: June 24, 2003, 10:18:29
 Job time : 20.9426 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Pur 20: June 24, 2003, 10:02:55 ; Search time 0.6979 seconds
(without alignments)
2059,300 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSHLPMTVPTKLPNGLEVEQ

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 5%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2538	99.5	519	1	DHA4_YEAST	P46367	saccharomyc
2	1695	66.5	519	1	DHA5_YEAST	P40047	saccharomyc
3	1455	57.1	497	1	DHAL_ASPERG	P41751	aspergillus
4	1452	56.9	499	1	DHA6_YEAST	P54115	saccharomyc
5	1412.5	55.4	497	1	DHAL_EMENI	P08157	emerella
6	1399	54.9	496	1	DHAL_CLAHE	P40108	cladosporiu
7	1362.5	53.4	495	1	DHAL_ALITAL	P42041	alternaria
8	1275.5	50.0	519	1	DHAM_MOUSE	P47734	mus musculu
9	1274.5	50.0	519	1	DHAM_PAT	P11894	rattus norv
10	1261.5	49.5	500	1	DHAM_MESAU	P11778	mesocricetu
11	1253.5	49.2	517	1	DHAM_HUMAN	P05041	homo sapien
12	1248.5	49.0	517	1	DHA5_HUMAN	P30837	homo sapien
13	1244.5	48.8	499	1	DHAS_CHICK	O33344	gallus gall
14	1240.5	48.6	520	1	DHAM_BOVIN	P20000	bos taurus
15	1230.5	48.2	499	1	DHA2_HUMAN	O94788	homo sapien
16	1223.5	48.0	499	1	DHA2_PAT	O63639	rattus norv
17	1222.5	47.9	499	1	DHA2_MOUSE	O62148	mus musculu
18	1220.5	47.9	500	1	DHAM_HORSE	P12762	equus cabal
19	1219	47.8	500	1	DHAL_AGART	O74187	ajaticus li
20	1218.5	47.8	512	1	DHA6_HUMAN	P47895	homo sapien
21	1212.5	47.5	504	1	DHAL_CHICK	P27453	gallus gall
22	1192.5	46.8	497	1	DHAL_FROBO	O57643	echinlatus
23	1182.5	46.4	500	1	DHAL_HUMAN	P09352	homo sapien
24	1181.5	46.3	500	1	DHAL_SHEEP	P51977	ovis aries
25	1179.5	46.3	500	1	DHAL_PAT	P51647	rattus norv
26	1176.5	46.1	500	1	DHAL_HORSE	P15437	equus cabal
27	1175.5	46.1	500	1	DHAL_MOUSE	P24549	mus musculu
28	1174.5	46.1	500	1	DHAL_BOVIN	P48644	bos taurus
29	1169.5	45.9	500	1	DHAL_PAT	P13601	rattus norv
30	1155.5	45.3	501	1	DHAM_MACPP	Q29430	macroscelid
31	1149.5	45.1	501	1	DHA6_FLEED	Q28439	elephantulu
32	1141	44.7	498	1	DHAM_LEITA	Q25417	leishmania
33	1101.5	43.2	495	1	CPOM_OCTDO	P30841	octopus dof

34	1099	44.1	902	1	FTDH_PAT	P28037	rattus norv
35	1095	42.9	494	1	CPOM_OMMSL	P40642	ommatreph
36	1092	42.8	506	1	DHA2_HUMAN	P47371	saccharomyc
37	1067	41.8	902	1	FTDH_HUMAN	O75891	homo sapien
38	1057	41.5	506	1	DHA3_YEAST	P54114	saccharomyc
39	1034.5	40.6	490	1	DHAB_BAGSU	P71015	bacillus su
40	1006	39.5	511	1	DHAY_YEAST	P42872	saccharomyc
41	948.5	37.2	499	1	FEAR_ECOLI	P80668	escherichia
42	942.5	37.0	487	1	DHAB_EHIME	P54232	rhizobium m
43	936	36.7	497	1	DHAB_STIOL	P17202	spinacia ol
44	930	36.5	512	1	ALDB_ECOLI	P37685	escherichia
45	919.5	36.1	506	1	DHAI_VIRCH	P23249	vibrio chol

ALIGNMENTS

RESULT 1				
DHA4_YEAST				
ID	DHA4_YEAST	STANDARD:	PRT:	519 AA.
AC	P46367: Q08898;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Potassium-activated aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (K(+)-activated acetaldehyde dehydrogenase) (K(+)-ACDH).			
GN	ALPD OP ALPD OP ALPD OP YOP374W OP 06730.			
OS	Saccharomyces cerevisiae (Baker's Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
KP	SEQUENCE FROM N.A.			
RA	Deliuss H., Robling U., Hofmann B.;			
RI	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 25-65.			
RX	MEDLINE=91113163; PubMed=1989592;			
RA	Chalmers P.M., Keen J.N., Fowson C.A.;			
RT	"Comparison of benzyl alcohol dehydrogenases and benzaldehyde dehydrogenases from the benzyl alcohol and mandelate pathways in <i>Actinobacter carboxticus</i> and from the TOL-plasmid-encoded toluene pathway in <i>Pseudomonas putida</i> . N-terminal amino acid sequences, amino acid compositions and immunological cross-reactions.";			
RL	Biochem. J. 273:99-107(1991).			
RP	SEQUENCE OF 25-34 AND 378-386.			
FX	MEDLINE=92295378; PubMed=9150920;			
PA	Larsson T., Norbeck J., Karlsson H., Karlsson K.-A., Blomberg A.;			
PT	"Identification of two dimensional gel electrophoresis resolved yeast proteins by matrix-assisted laser desorption/ionization mass spectrometry.";			
RL	Electrophoresis 18:418-423(1997).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND CHARACTERIZATION.			
FX	MEDLINE=93340498; PubMed=9275847;			
PA	Tessier W.H., Meaden P.G., Dickinson F.M., Midgley M.;			
PT	"Identification and disruption of the gene encoding the F(1)-activated acetaldehyde dehydrogenase of <i>Saccharomyces cerevisiae</i> .";			
FL	FEMS Microbiol. Lett. 164:29-34(1998)			
OC	!! CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> an acid + NADH.			
OC	!! SUBCELLULAR LOCATION: Mitochondrial matrix.			
OC	!! SIMILARITY: RELATES TO THE ALDH4 DEHYDROGENASES FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/annouco/ or send an email to license@isb-sib.ch)			
CC	EMBL; 275282; CAA99705.1; -			

```

DR HSSP: P05091; ICW4.
DR SCD: S0005901; ALD4.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS006070; ALDEHYDE_DEHYDR_GYS; 1.
DR PROSITE: PS006087; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transl. peptide.
FT TRANSIT 1 24 MITOCHONDRION.
FT CHAIN 25 519 POTASSIUM-ACTIVATED ALDEHYDE
DEHYDROGENASE.
FT NP_BIND 268 273 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 290 290 POTENTIAL.
FT ACT_SITE 324 324 POTENTIAL.
FT ACT_SITE 51 51 N -> NN (IN REF. 2).
FT CONFLICT 63 63 E -> V (IN REF. 2).
SQ SEQUENCE 519 AA: 56723 MW: 6709444EAS: FQ48E CFF64.

Query Match 99.5% Score 2538; DB 1; Length 519,
Best Local Similarity 99.8%; Pred. No. 2.6e-158;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 SHLPMTVPKLPNGLEYEQTLGFTLNKFKVPSKQNTFEVINPSTEEIECHIIYEGREDDV 61
DQ 25 SHLPMTVPKLPNGLEYEQTLGFTLNKFKVPSKQNTFEVINPSTEEIECHIIYEGREDDV 84
QY 62 EEAQAADRAFSNCSNNGIDIPDRGKALYRLAELIEQDKDVLIASLETDLNGKAISSSRGD 121
DQ 85 EEAQAADRAFSNCSNNGIDIPDRGKALYRLAELIEQDKDVLIASLETDLNGKAISSSRGD 144
QY 122 VDLVINYLYKSSAGFAKIDGPMIDTGTTHFSYTKPGIPLVCGGIIIPWNPFLIMAWKAP 181
DQ 145 VDLVINYLYKSSAGFAKIDGPMIDTGTTHFSYTKPGIPLVCGGIIIPWNPFLIMAWKAP 204
QY 182 ALVTGNTVWLKTAESTPLSALYSKYIIPGAPPGVINIVSGEKGIVVEALTNHPRIKKV 241
DQ 205 ALVTGNTVWLKTAESTPLSALYSKYIIPGAPPGVINIVSGEKGIVVEALTNHPRIKKV 264
QY 242 AFTGSTATGTHIYCSAAGLKKVITLGGCKSNIVFADAEKKAVONITLGIYNSGEVC 301
DQ 265 AFTGSTATGTHIYCSAAGLKKVITLGGCKSNIVFADAEKKAVONITLGIYNSGEVC 324
QY 302 CAGSKRVVRESIVDKTLEEFKAASESIKVGDDPDESTFGCAQTSQWLNLKILKYVDIGKN 361
DQ 325 CAGSKRVVRESIVDKTLEEFKAASESIKVGDDPDESTFGCAQTSQWLNLKILKYVDIGKN 384
QY 362 EGATLITGGERLGSKGYFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADEVINMAND 421
DQ 385 EGATLITGGERLGSKGYFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADEVINMAND 444
QY 422 SEYGLAAGIHTSNITALKVADRVNAGTYWINTYNDFFHVAFFPGFNAGSLGRENVDAL 481
DQ 445 SEYGLAAGIHTSNITALKVADRVNAGTYWINTYNDFFHVAFFPGFNAGSLGRENVDAL 504
QY 482 QNYIQVKAQVAKLID 495
DQ 505 QNYIQVKAQVAKLID 518

RESULT 2
DHA5_YEAST STANDARD; PRT; 519 AA.
AC P40047;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3)
GN ALD5 OR ALD5 OR ALD3 OR YER073W
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 11
RP SEQUENCE FROM N.A.

KX MEDLINE-97212814; PubMed-9059641;
KA Wang X., Bai Y., Ni L., Weiner H.;
KT "Saccharomyces cerevisiae aldehyde dehydrogenases: Identification and
expression.";
PL Adv. Exp. Med. Biol. 414:277-280(1997).
FN [2]
KP SEQUENCE FROM N.A.
KC STRAIN-S288C / AB972;
KA Dietrich F.S., Mulligan J.T., Hennessy K.M., Allen E., Aranda R.,
PA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
PA Chung F., Duncan M., Guzman E., Hartzell G., Hunkeler Smith S.,
PA Hyman R., Kayser A., Komp C., Lashkari D., Law H., Lin D.,
PA Mosedale D., Nakahara K., Namath A., Norton R., Oefner P., O'Leary
PA Petel F.X., Roberts D., Schi P., Schramm S., Shenton T., Smith V.,
KA Taylor P., Wei Y., Yellon M., Rotstein D., Davis R.W.;
PL Submitted (DEC-1994) to the EMBL/GenBank/Trinac databases.
FN [3]
KP CHARACTERIZATION.
KP MEDLINE-98132377; PubMed-9474045;
KA Wang X., Mann C.J., Bai Y., Ni L., Weiner H.;
KT "Molecular cloning, characterization, and potential roles of cytosolic
and mitochondrial aldehyde dehydrogenases in ethanol metabolism in
Saccharomyces cerevisiae.";
KL J. Bacteriol. 180:822-830(1998).
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix (potential).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

This SWISS PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed using any and/or commercial
entities requires a license agreement (See: http://www.isb-sib.ch/announcements/
or send an email to license@isb-sib.ch).

DQ EMBL: U56605; AAB01220.1;
DR EMBL: U18814; AAB64612.1;
DR HSSP: P51977; IHX5.
DR SCD: S0000875; ALD5.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS006070; ALDEHYDE_DEHYDR_GYS; 1.
DR PROSITE: PS006087; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transl. peptide.
FT TRANSIT 1 23 MITOCHONDRION (potential).
FT CHAIN 24 519 ALDEHYDE DEHYDROGENASE.
FT NP_BIND 265 270 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 287 287 BY SIMILARITY.
FT ACT_SITE 321 321 BY SIMILARITY.
FT CONFLICT 48 48 T -> I (IN REF. 2).
FT CONFLICT 90 90 L -> A (IN REF. 2).
FT CONFLICT 93 103 LUKRSVYCHAG -> APTKRSIVVEP (IN REF. 2)
FT CONFLICT 410 410 E -> G (IN REF. 2).
SQ SEQUENCE 519 AA: 56552 MW: 5034354379773 CFF64.

Query Match 66.5% Score 1695; DB 1; Length 519;
Best Local Similarity 63.7%; Pred. No. 2.4e-103;
Matches 316; Conservative 80; Mismatches 94; Indels 6; Gaps 2;

QY 2 SHLPMTVPKLPNGLEYEQTLGFTLNKFKVPSKQNTFEVINPSTEEIECHIIYEGREDDV 61
DQ 24 SHLPMTVPKLPNGLEYEQTLGFTLNKFKVPSKQNTFEVINPSTEEIECHIIYEGREDDV 84
QY 62 EEAQAADRAFSNCSNNGIDIPDRGKALYRLAELIEQDKDVLIASLETDLNGKAISSSRGD 119
DQ 84 EEAQAADRAFSNCSNNGIDIPDRGKALYRLAELIEQDKDVLIASLETDLNGKAISSSRGD 149
QY 120 VDLVINYLYKSSAGFAKIDGPMIDTGTTHFSYTKPGIPLVCGGIIIPWNPFLIMAWKAP 179
DQ 140 VDLVINYLYKSSAGFAKIDGPMIDTGTTHFSYTKPGIPLVCGGIIIPWNPFLIMAWKAP 199

```



```

QY 180 APALVTGNTVTLKTAESTPLSALYVSKYIPQAGIPPGVIVNIVSGSKIVVEAFITNHPKIK 249
DB 200 GPALATGNTVTLKPAETTHLSALPASQAQAGIPAGVWNLISGSGVVGSKISAIHPDVK 259
QY 240 KVAFTGSTATGPHIYQSAAGIKKVTIELGSKSNIVFAIAIKKAVQNIILGLIYNSRE 269
DB 260 KIAPTGSTATGPHIMKVAAGVKKVTLIELGSKSNIVFAIADLIKAVKNIAPGTFYNSGE 319
QY 300 VCCAGSPVYVESIYDKELBEFKKAESIKVGDPEDESIHQAGVISMJLNKILKYVDIG 359
DB 320 VCCAGSRIVQDTVYEEVLQKDKYTESLKVGDPEDEEVFGAGTSDKQJRKILDYDVA 379
QY 360 KNGATLITGGERLASKGYFIKPTVFGVSKEDMBIVKEEIPGVPVTVTKFKSADEVINMA 419
DB 380 KSGAPVITGGAPHSKGYFVKPTVFAVDKEDMBIVKEEIPGVPVTVTKFKSADEVINMA 439
QY 420 NDSYVGLAAGIHTSNITALKVADRNVAGTVMINTYNDPHHAFVFGFGNAGIGREMSVD 479
DB 440 NDSYVGLAAGIHTSNITALKVADRNVAGTVMINTYNDPHHAFVFGFGNAGIGREMSVD 499
QY 480 ALQNYLVQKAVPAKL 495
DB 500 ALSNTQTQKSVRIAD 515

RESULT 3
DHAL_ASPNG
ID DHAL_ASPNG STANDARD: PRT: 497 AA.
AC P41751;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ALdehyde dehydrogenase (EC 1.2.1.3) (ALDH1).
GN ALDH1.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N A
PX MEDLINE=90108706; PubMed=2606357;
RA O'Connell M J, Kelly J M.;
FT "Physical characterization of the aldehyde dehydrogenase-encoding
FT gene of Aspergillus niger";
RL Gene 84:173-180(1989)
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright © 1995 and is reproduced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M32351; AAA87596.1;
CC HSSP; P05091; LCM3.
CC InterPro: IPR002086; Aldehyde_dehydr
CC Pfam: PF00171; aldehyd; 1.
CC PROSITE: PS00070; ALDEHYDE_DEHYDR_GYS; 1
CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
CC Oxidoreductase; NAD.
CC NP_BIND; 242 247 NAD (ADP PART) (BY SIMILARITY)
CC ACT_SITE; 265 265 BY SIMILARITY.
CC ACT_SITE; 299 299 BY SIMILARITY.
CC SEQUENCE; 497 AA; 53806 MW; 47CA53F08A508A7654;
Query Match 57 18; Score 1455; DR 1; length 497;
Best Local Similarity 56.08; Pred. NO.1e-87;
Matches 277; Conservative 83; Mismatches 131; Indels 4; Gaps 3;

```

RESULT 4

DHA6_YEAST

```

ID DHA6_YEAST STANDARD: PPT: 499 AA.
AC P54115; Q02783;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Magnesium-activated aldehyde dehydrogenase, cytosolic (EC 1.2.1.3)
DE (Mg(2+)-activated acetaldehyde dehydrogenase) (Mg(2+)-ACDH).
GN ALDH6 OR ALDH1 OR YPL061W OR LPE9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N A, AND CHARACTERIZATION.
PX STPAIN-DBY939;
PX MEDLINE=9813277; PubMed=9473035;
RA Wang X, Mann C J, Bai Y, Ni L, Weiner H.;
RT "Molecular cloning, characterization, and potential roles of cytosolic
RT and mitochondrial aldehyde dehydrogenases in ethanol metabolism in
RT Saccharomyces cerevisiae.";
FL 1. Nucleic Acids Res 22:402-408(1994).
RN [2]
RP SEQUENCE FROM N A.
PX STRAIN=S288C / AB972;
PX MEDLINE=97313271; PubMed=9169875;
RA Russ J H, Strimling F K, Ahmed A, Albrecht K, Allen F, Ansorge W,
RA Araujo R, Aparicio A, Barrell B G, Badcock K, Benes V,
RA Botstein D, Bowman S, Bruckner M, Carpenter J, Cherry J M,
RA Chung E, Churcher M, Cosserat F, Davis K, Davis R W,
RA Dietrich F S, Delius H, Dipaolo T, Dubois E, Duesterhoeft A,
RA Duncan M, Floeth M, Fortin N, Friese J P, Fritz C, Goffeau A,
RA Hall J, Hebling U, Heumann K, Hilbert H, Hillier L,
RA Hunicke-Smith S, Hyman R, Johnston M, Kalman S, Klein K,
RA Komp C, Kurdi G, Lashkari D, Low H, Lin A, Lin D, Louis F J,

```


QY 1 MSILPMTVPKLENGLEVEOPTGLFINNKFVPSKQNKTEFVINSTEETEECHIEVEDEED 60
 DB 1 MSDFLTITETTPV---IKYELQGLGFINNEFVKGVEGKTFQVINNEKVFITSVHEATEKD 57
 QY 61 VFFAVCAADPAFENFSGWNGTIPIDPKALFYPIAFITFCKKIVTASITFLLNGKALSSSG 120
 DB 58 VDVAVAAAPAF-FEPWPCVIVSEPGIILNKLALUMERULILAAIESLNGKAFIMAKV 116
 QY 121 DVLIVINLYKSAGFADKIDGMIDITGTHPSYTKRQPLGVCQGIIPWNPPLLMWAKIA 180
 DB 117 DLANSIGTQRYTAAWADRIHQITPILINPEHIIYTHPEFVSGIIPWNPPLLMWSKIG 176
 QY 181 PALVTGNTVLTABSTPLSALYSKYIPOAGIPGIVINISVSGKIVVEATINHPKIKK 240
 DB 177 PAVAGNTVLTAGTQPLSALYAAKLIKFAFFAGVINISVSGRTAGAAALSHMDIDK 236
 QY 241 VAFGTSTAGRHIYOSAA-AGLKVVTELGKSNIPVADAEALKKAVONITLIGIYNSGE 299
 DB 237 VAFGTSTIGVPTIIQAAAKSNLKVVTELGKSNIPVADAEALKKAVONITLIGIYNSGE 296
 QY 300 VCCAGSRVVEESIYDKTEEFKKAESIKYGDPEDESTFOGATQTSOMOLNKLKLYVDIG 359
 DB 297 CCAGSPILVTEIYDKTEEFKKAESIKYGDPEDESTFOGATQTSOMOLNKLKLYVDIG 356
 QY 360 KNEGATLITGGERLGSKYFIKPTVFGDKEDMRIVKEEIPGVPVTVTKFSADEVINMA 419
 DB 357 KKAGATVATGGDRHNGEYFQPTVFTDVTSDMKIAQEEIPGVPVTVTKFSADEVINMA 416
 QY 420 NDSYGLAAGIHTSNITALKVADRVNAGTVWINTYDFHVAFFGFGNAGSLGRENMSVD 479
 DB 417 NSTDYGAAAVHTKNVTAIRVSNALKAGTVWINTYDFHVAFFGFGNAGSLGRENMSVD 476
 QY 480 ALONYLQVKAERAKL 494
 DB 477 ALENYTOIKTVHYRL 491

RESULT 6
 ID DHAL_CLAHE STANDARD, PRT: 496 AA
 AC P40108;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH) (Allergen Cla h 3) (Cla h
 II).
 GN CLAH3.
 OS Cladosporium herbarum
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
 OC Dothideomycetes et Chaetothyrionomycetes incertae sedis;
 OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
 OX NCBI_TaxID=29918;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=240202-Berlin;
 RX MEDLINE=95206305; PubMed=7898496;
 RA Achatz G., Oberkofler H., Lechenauer E., Simon B., Unger A.,
 RA Kandlauer P., Ehner C., Prillinger H., Kraft P., Breitenbach M.;
 RT "Molecular cloning of major and minor allergens of Alternaria
 RT alternata and Cladosporium herbarum."
 RL Mol. Immunol. 32:213-227(1995).
 CC -/- CATALYTIC ACTIVITY An aldehyde + NAD(+) -> H(2)O - an acid + NADH.
 CC -/- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -/- SIMILARITY: RELIGIONS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -/- CAUTION: Two distinct proteins have been termed allergen
 CC cla h 3

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to info@isb-sib.ch).

 CC EMBL: X78228; CAAS5072.1; -
 DR HSSP: P05091; ICW3.
 LF InterPro: IPR002046; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR KOSMIE: PS00070; ALDEHYDE_DEHYDRO_CYS; 1.
 DR PROSITE: PS00687; ALPHAHYDRODEHYDROGLU; 1.
 KW Oxidoreductase; NAD; Allergen.
 FT AC1_SITE 253 263 BY SIMILARITY.
 FT AC1_SITE 296 296 BY SIMILARITY.
 SQ SEQUENCE 496 AA: 53397 MW: 43376606.1679908 CFC64;

Query Match 54.98; Score 1399; DB 1; Length 496;
 Best Local Similarity 53.84; Pred No. 4 aa-84;
 Matches 263; Conservative 89; Mismatches 145; Indels 2; Caps 2.
 QY 7 IVPLKPLNGLEYEQVIGLFINNKFVPSKQNKTEFVINSTEETEECHIEVEDEEDVLA 66
 DB 3 SVQLETSHSGKYEOTGLFINNKFVPSKQNKTEFVINSTEETEECHIEVEDEEDVLA 62
 QY 67 AADAFASNGSWNGIILUPIDPKALFYPIAFITFCKKIVTASITFLLNGKALSSSG 126
 DB 63 AADAFASNGSWNGIILUPIDPKALFYPIAFITFCKKIVTASITFLLNGKALSSSG 121
 QY 127 NYLKSSAFPAKLIKRMIDTGTHTSVTKRQPLGVCQGIIPWNPPLLMWAKIA 186
 DB 122 GCLRYGGWADKITGKVIDITPTDPTFNVYKKEPIGVCRSDHSLPLLMWAKI 181
 QY 187 NTWVLTAEPTISAIYVSKYIPGAGIPGIVINISVSGKIVVEATINHPKIKKVA 246
 DB 182 NTWVLTAEPTISAIYVSKYIPGAGIPGIVINISVSGKIVVEATINHPKIKKVA 241
 QY 247 TATGPHIYOSAA-AGLKVVTELGKSNIPVADAEALKKAVONITLIGIYNSGE 305
 DB 242 TVGRTILKAASSNLKVVTELGKSNIPVADAEALKKAVONITLIGIYNSGE 301
 QY 306 RVVVEESIYDKTEEFKKAESIKYGDPEDESTFOGATQTSOMOLNKLKLYVDIG 365
 DB 302 RVVVEESIYDKTEEFKKAESIKYGDPEDESTFOGATQTSOMOLNKLKLYVDIG 361
 QY 366 LITGERLGSKYFIKPTVFGDKEDMRIVKEEIPGVPVTVTKFSADEVINMA 425
 DB 362 VETGSRKSGKYFETPTFISNVTEDMKIVKEEIPGVPVTVTKFSADEVINMA 421
 QY 426 LAAGTHTSNITALKVADRVNAGTVWINTYDFHVAFFGFGNAGSLGRENMSVD 485
 DB 422 LAAGTHTSNITALKVADRVNAGTVWINTYDFHVAFFGFGNAGSLGRENMSVD 481
 QY 496 QVKAERAKL 494
 DB 482 QTKVTSIRL 490

RESULT 7
 ID DHAL_ALIAL STANDARD, PRT: 495 AA.
 AC P42041;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 LE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH) (Allergen Alt a 10) (Alt a
 X).
 GN ALTA10.
 OS Alternaria alternata.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
 OX NCBI_TaxID=5599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=08-0203-Berlin;
 RX MEDLINE=95206305; PubMed=7898496;

QY 300 VCCAGSRVYVESIYDKFIEEPKAASEIKVGDPPDESTFQCAQTSOMOLNKLKYVDIG 359
 DB 318 CCCASPTFQEDIDYEFVSVARPAKSPVNPFTSKTEQSPQVDFTKFKILGYINTG 377
 QY 360 KNECATLITGGERLGSKGYYIKPTFGDYKEMPIVKEELFGPVTVTKFSADENVMA 419
 DB 378 KQELAKILCCGGLAARDGYPIQTVFGDVQDGMNTAKKEELFGPVWMLKFKTIEEVGGA 437
 QY 420 NDSYGLAAGHTSNINTALKVADRVNAGTIVINTYNDHFHVPFGFNASGLGEMSVSD 479
 DB 438 NNSTYGLAAAVFKDLIDKANYLSQALQAGIIVWNCYDVPFGAQSPTFGYKMGSGRELGEY 497
 QY 480 ALONYLOVKAVRAKL 494
 DB 498 GLQAYTEKTVTKV 512

RESULT 12

DHA5_HUMAN
 ID DHA5_HUMAN STANDARD; PRT; 517 AA.
 AC P30837.
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase X, mitochondrial precursor (PC 1 2 3) (ALDH class 2).
 GN ALDH1B1 OR ALDH5 OR ALDHX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=91286241; PubMed=2061311;
 RA Hsu L.C., Chang W.-C.;
 RT "Cloning and characterization of a new functional human aldehyde dehydrogenase gene."
 RL J. Biol. Chem 266:12267-12265(1991)
 RN [2]
 RP VARIANTS VAL-86 AND ARG-107.
 RX MEDLINE=94063858; PubMed=8244338;
 RA Sherman D., Hsu V., Hsu L.C., Peters T.J., Yoshida A.;
 RT "Diverse polymorphism within a short coding region of the human aldehyde dehydrogenase-5 (ALDH5) gene."
 RL Hum. Genet. 92:477-480(1993).
 CC FUNCTION: ALDH5 PLAY A MAJOR ROLE IN THE DETOXIFICATION OF ALCOHOL-DERIVED ACETALDEHYDE. THEY ARE INVOLVED IN THE METABOLISM OF CORTICOSTEROIDS, BIOGENIC AMINES, NEUROTRANSMITTERS, AND LIPID PEROXIDATION.
 CC CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC PATHWAY: Ethanol utilization; second step.
 CC SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC TISSUE SPECIFICITY: LIVER, TESTIS AND IN A LESSER EXTENT IN BRAIN.
 CC SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/annouce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M63967; AAA96830.1; -
 CC HSP: P20000; IAG8
 CC Genew: HGNC:407, ALDH1B1.
 CC MIN: 100670; -
 CC InterPro: IPR002086; Aldehyde-dehydr.
 CC Pfam: PF00171; aldedh; 1.
 CC PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.

DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1
 KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism.
 FT TRANSIT 1 17 MITOCHONDRION (POTENTIAL)
 FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE X.
 FT NP_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 285 285 POTENTIAL.
 FT ACT_SITE 319 319 POTENTIAL.
 FT VARIANT 86 86 A -> V (IN ALDH1B1*2).
 FT 107 107 /FTID VAR_002257.
 FT 107 107 L -> R (IN ALDH1B1*3).
 FT 107 107 /FTID VAR_002258.
 SQ SEQUENCE 517 AA; 57217 MW; 78A174AA687165 CRC64;

Query Match 49.0%; Score 1248.5; DB 1; Length 517;

Best local Similarity 50.8%; Pred. No. 3,2674;

Matches 248; Conservative 85; Mismatches 148; Indels 9; Gaps 6;

QY 9 P K L P N G L Y E Q P T G L F I N N K F V P S K N K T F V I N I S T E E E T C H I V E G R E I D V E A V A A 68
 DB 28 P I L N P D - I P Y N O - - - L F I N N E W O D A V S K K T F P V N I T G E V I G H V A E G I R A I V D R A V A A 84
 QY 69 L P A F S N S - W N G I D P I L P K A L Y P I A E L I E D I K R I V I A S I F T I N N K A I S S S D E - I V D I V I 126
 DB 84 P E A F L E S E P W P R M I A S E F G K L I N L A D I V E R G V I A S E F T I N N K K P E Q E S T A L D E R V I 144
 QY 127 N Y I K S S A G C A D K I P G R M I D P T H E S Y T K P Q P G W G G I I P W N P I L M W A N K I A I A L V I G 186
 DB 144 K V Y Y F A G W A D K W H G K T I P M H Q G H F T T R I E H V G V G Q I I P W N P I V M G W K L A I A L A I G 204
 QY 187 N T V L K T A R S T P L S A L Y S K Y I P Q A I P P T V I N I S C P K I V V E A I T N H P I F F V A F P S 246
 DB 204 N T V M K V A E Q T P L S A L Y S L I K E A G P P G V V N I I P G Y P T A - A A I A Q H M I V K V A F S 264
 QY 247 T A T E R H I Y A S A A - A S L K P V I L F E A K S N I V F A V A L K K A V N I I L I C I Y N S G E V A A 304
 DB 264 T E V G - H L Y K A A G S N L K V I T E L E A K S P S I V A I A D I A V I G - T H E A I T N M G V A G 322
 QY 305 S R Y V E E S I Y D K F I E F K A A S E S I K V G D P D E S T F Q C A Q T S O M O L N K L I K Y V D I G K N E G A 364
 DB 323 S P T T V E E S I Y N F L E E T V E K A K G F K V K N I T E - T T E - P - V I P T E P V E V I A T I G K N E A 382
 QY 365 T I L T G G E R I G S K G Y F I K P T V F G I V K E D K M I V K E E I P G V V T V I P K S A I E V I N M A N S E Y 424
 DB 384 K L I G G E R F E R G E F F I K P T V G - V Q D M P I A F E E I P G V P I P R E F P I E F V V F P A N N T K Y 442
 QY 425 G L A A G H T S N I N T A L K V A D R V N A G T V I N T Y N D F H H A V P F G G F N A S G L G E M S V I A L O N Y 484
 DB 443 C L A A A V P T K D L K A M Y F T U A L Q N C I V W N Y N I V I C H T P E G E K I S E N G P R E L G E L K A Y 502
 QY 485 L Q V K A V R A K L 494
 DB 503 T E V K T V T I K V 512

RESULT 13

DHA5_CHICK
 ID DHA5_CHICK STANDARD; PRT; 499 AA.
 AC 093344;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aldehyde dehydrogenase 1A2 (EC 1.2.1.3) (retinaldehyde-specific dehydrogenase type 2) (PALDH1(1)) (PALDH-2).
 GN ALDH1A2 OR PALDH2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9839476; PubMed=9727494;
 RA Sockanathan S., Jessell T.M.;


```

FT VARIANT 409 409 E -> V
SQ SEQUENCE 497 AA: 54138 MW: 9C6CF62731A25BD CRC64:

Query Match
Best Local Similarity 56.1%; Score 1429.5; DB 3; Length 497,
Matches 273; Conservative 87; Mismatches 130; Indels 5; Gaps 3;

QY 1 MSHLPMTVPIKLNGLLEYEQPTCLFINNKFVPSKONKTFEEVINSPEEIECHITGREGD 60
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1 MSDFLTITETPV IKYEUGLGFINNEFVGVEKTFQVTPNSNEKVIISVHEAIEKD 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 61 VEFVQAADRAFSNGSNGWGDIDRCKALYRLAELEGGKLVIASTIELDNGKALISSKG 120
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 58 VDVAVAAARAAP-EGPWRQVTPSEKGIINKLADUMERDIDTAALESILDNCKAFIMAKV 116
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 121 DVDLVINYLKSSAGPAIKTDCPMITDTGPTHTSVYTKRQVGVGGTTPWNPFLIMMAWKIA 180
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 117 DLANSIGCLRYAGWADKIHGQITIDNPETLTITRHPVGVGGQIPWNPFLIMMSWKIG 176
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 181 PALVTGNTVVKTAESTPLSALYVSKYIPQAGIPGCVINIVSGKIVVEAIIHHPKIKK 240
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 177 PAVAAGNTVVKTAETPTLSALYAAKLIKPAEPFACVINVISRGFTAGAAISSHMDIDK 236
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 241 VAFPGSTATGRHIYQSAA-AGLKVKVLEIGCKSPNIVFADAEKKAVGNIIILGIYVNSGE 299
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 237 VAFPGSTLVGRTILQAAKSNLKKVLELGGKSPNIVFDADIDNAISWANFGIFENHGQ 296
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 300 VCCAGSVVVEESIYDKETEEFKAASESILKVGDPDESTEGAGTSOMCLNKLKLYVDIG 359
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 297 CCAGSPHIVVGGIYDKFVARKERAKKNGNPFEDTFQGVQSLOFDTIMFY-NHG 356
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 360 KNEGATLITGCKELGSKGYFIKPTVGDYKEDMRIVKEELFGPVVTVTKFSADDEVINMA 419
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 357 KKAGATVATGDRHNGEYFIQPTVPTDVTSDMKTAQEFIFGPPVTVTKFKDEAEAIKIG 416
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 420 NDSYGLAAGIHTSNINTALKVADRVNAGTVINYNDEHFAVPPGCGNASCLPFMSVD 479
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 417 NSTDYGIAAAVHTKNVNTAIRVSNALKAGTVINNNYMSIQAPGPGKQSGLGRFGSY 476
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 480 ALQNYLQVKAVRAKL 494
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 477 ALQNYTQIKTVHYRL 491
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 6
Q9C1Q4 PRELIMINARY; PRT: 497 AA.
AC Q9C1Q4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Aldehyde dehydrogenase ALDH57.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes.
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Flippin M., Mathion M., Cirpus I., Panopze C., Fellenberg B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AF260125; AAK18074.1;
DR HSSP: P05091; 1CW3.
DR InterPro: IPR002086; Aldehyde-dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDPHYDE_DEHYDR_CYS; 1
DR PROSITE: PS00687; ALDPHYDE_DEHYDR_G10; 1.
KW oxidoreductase.

FT VARIANT 338 338 S -> G
SQ SEQUENCE 497 AA: 54168 MW: 54F425A84B4064875 CRC64:

Query Match
Best Local Similarity 54.9%; Score 1424.5; DB 3; Length 497,
Matches 272; Conservative 87; Mismatches 131; Indels 5; Gaps 3;

QY 1 MSHLPMTVPIKLNGLLEYEQPTCLFINNKFVPSKONKTFEEVINSPEEIECHITGREGD 60
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1 MSDFLTITETPV IKYEUGLGFINNEFVGVEKTFQVTPNSNEKVIISVHEAIEKD 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 61 VEFVQAADRAFSNGSNGWGDIDRCKALYRLAELEGGKLVIASTIELDNGKALISSKG 120
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 58 VDVAVAAARAAP-EGPWRQVTPSEKGIINKLADUMERDIDTAALESILDNCKAFIMAKV 116
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 121 DVDLVINYLKSSAGPAIKTDCPMITDTGPTHTSVYTKRQVGVGGTTPWNPFLIMMAWKIA 180
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 117 DLANSIGCLRYAGWADKIHGQITIDNPETLTITRHPVGVGGQIPWNPFLIMMSWKIG 176
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 181 PALVTGNTVVKTAESTPLSALYVSKYIPQAGIPGCVINIVSGKIVVEAIIHHPKIKK 240
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 177 PAVAAGNTVVKTAETPTLSALYAAKLIKPAEPFACVINVISRGFTAGAAISSHMDIDK 236
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 241 VAFPGSTATGRHIYQSAA-AGLKVKVLEIGCKSPNIVFADAEKKAVGNIIILGIYVNSGE 299
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 237 VAFPGSTLVGRTILQAAKSNLKKVLELGGKSPNIVFDADIDNAISWANFGIFENHGQ 296
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 300 VCCAGSVVVEESIYDKETEEFKAASESILKVGDPDESTEGAGTSOMCLNKLKLYVDIG 359
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 297 CCAGSPHIVVGGIYDKFVARKERAKKNGNPFEDTFQGVQSLOFDTIMFY-NHG 356
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 360 KNEGATLITGCKELGSKGYFIKPTVGDYKEDMRIVKEELFGPVVTVTKFSADDEVINMA 419
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 357 KKAGATVATGDRHNGEYFIQPTVPTDVTSDMKTAQEFIFGPPVTVTKFKDEAEAIKIG 416
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 420 NDSYGLAAGIHTSNINTALKVADRVNAGTVINYNDEHFAVPPGCGNASCLPFMSVD 479
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 417 NSTDYGIAAAVHTKNVNTAIRVSNALKAGTVINNNYMSIQAPGPGKQSGLGRFGSY 476
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 480 ALQNYLQVKAVRAKL 494
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 477 ALQNYTQIKTVHYRL 491
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 7
Q14293 PRELIMINARY; PRT: 503 AA.
AC Q14293;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE Probable aldehyde dehydrogenase (EC 1.2.1.4) (Alb000).
GN SPAC9E9.09C.
OS Schizosaccharomyces pombe (Pission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes.
OC Schizosaccharomycetales; Schizosaccharomycetaceae.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX SPEAN 972;
RA McDougal R., Bartell B.G., Kallandrom M.A., Wood V.;
RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H2O -> ALD + NADH
CC -1- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: Z994262; CAB16407.1;
DR HSSP: P05091; 1CW4.
DR InterPro: IPR002086; Aldehyde-dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_G10; 1.
KW oxidoreductase; NAD.

```



```

Db 299 VKKVTELEGGKSPNIFAADADLEAVRQSHHALFFNOGQCSAGSKTFVEGKIYDEFVAK 358
QY 321 FKAASESTKWDDEPDESTFOCAOTSQMLNKLKYYVDIGKNEDALITIGSEGLGSKGYF1 380
Db 359 AKELVETVIGDPFDENTTGGPDIDQSVETIMKYIESGKKEGAQLVTGGVKGHGQGYFV 418
QY 381 KPTVFGDVKEMIRIVKEEIPGVVTVTKPKSADEVIMNANDSEYGLAAGIHTSNTALK 440
Db 419 KPTIFANVDMOKTIAOEIPGVMIIVIRFDSMEELIEKANNTIYGLAAGVTVNDLKNALQ 478
QY 441 VADRVNAGTVWINTYDNFHHAVPFGGFGNAGLGKREMSVDALQNYLOVKAVRAKL 494
Db 479 VANTIRAGSVVANCYDVFDAAPPFGGKQSGTGRGELGEYGLAAYTEVKTVTIKV 532

RESULT 13
Q9BV45 PRELIMINARY; PRT: 517 AA.
ID Q9BV45;
AC Q9BV45;
DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE Similar to aldehyde dehydrogenase 5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata, Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=FFV;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DF EMBL: AC001619; AAH01619.1;
DR HSSP: P05091; ICW3.
DR InterPro: IPR02086; Aldehyde_dehydr.
DR Pfam: PF00171; alcdch.1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 517 AA; 57238 MW; A628E448D1E8689D CRC64;

```

```

Query Match 49.0%; Score 1249.5; DB 4; Length 517;
Best Local Similarity 50.8%; Pred. NO. 1.2e-76;
Matches 248; Conservative 85; Mismatches 148; Indels 9; Gaps 6;
QY 9 PIKPLNGLEYEQTGLFINNKFVPSKQKTFEVIINPSTEEETCHIEGREDDVEAVQAA 68
Db 28 PILNFD-IPYQ...LFINMQDAVSKKTFEVIINPSTEEETCHIEGREDDVEAVQAA 83
QY 69 DRAFSNKS-WNGIDPIDRGKALYLAELIEQDKVIASIFTLONGKATSSSPG-DVRLVI 126
Db 84 REAFPGASPPMPDASPPGLINLADIVERPVIASIFTLONGKATSSSPG-DVRLVI 143
QY 127 NYLKSSAGFADKIDGRMDITGRHESYTKRQPLGVCGQIIPWNPFLIMANKIAIAFVIG 186
Db 144 KVVYTAGWADKKHGTIPMGCHPTTHHEPVGVGQIIPWNPFLIMANKIAIAFVIG 203
QY 187 NTVVLKTAESTPLSALYVSKYIPQAGTPPGVINTVSYRPGKIVVEATINHPKIKFVAFTGS 245
Db 204 NTVMKVAEOTPLSALYSLASLKEAGFPFGVNNITGGYPTAGAAIAGHMTVTKVAFTGS 263
QY 247 TATGRHIYQSA--AGLKKVTLIELGKSPNIFAALAEKLVAVONIILGIYNSHEVYAG 304
Db 264 TENG-HLIOKAAGSKNRVTLIELGKSPNIFAALAEKLVAVONIILGIYNSHEVYAG 322
QY 305 SRYVVEESIVDKPIEFKAASESTIKVGPFDESTFOCAOTSQMLNKLKYYVDIGKNEDAL 364
Db 323 SPTEVESLYNELEPTVEKAKQPKVGNPFELTPGSGPQVQKESPEFVLGIQLGREGCA 382
QY 365 TLTGGERLGSKGYFTKPTVFGGVDFMDPTVKKFTFPGVTVTKFSADEVIMNANDSEY 424
Db 383 KLLGGGPFGEERGFETKPTVFGGVQDDMKIAKEETFGPVQPLFKPKKTEEVVERANTRY 442

```

```

QY 425 GLAAGIHTSNTALKAVDRVNACTVWINTYDNFHHAVPFGGFGNAGLGKREMSVDALQNY 484
Db 443 GLAAAVTEHEDFAMETLAAAGIYVWNYINIVTHITFGPKSGNPFEGEGRKAY 502
QY 485 LOVKAVRAKL 494
Db 503 TEVKTVTIKV 512

RESULT 14
Q9FRX7 PRELIMINARY; PRT: 549 AA.
ID Q9FRX7;
AC Q9FRX7;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Aldehyde dehydrogenase ALDH2B.
GN ALDH2B.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE; TISSUE=LEAVES OF SEEDLINGS;
RA Nakazono M.; Tsuji H.;
KT "Rice mitochondrial aldehyde dehydrogenase ALDH2B."
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DF EMBL: AB044537; BAB19052.1;
DR HSSP: P05091; ICW3.
DR InterPro: IPR062386; Aldehyde_dehydr.
DR Pfam: PF00171; alcdch.1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 549 AA; 59306 MW; 2B34C2B881FEB6BD CRC64;

```

```

Query Match 48.9%; Score 1247.5; DB 10; Length 549;
Best Local Similarity 49.8%; Pred. NO. 1.8e-76;
Matches 241; Conservative 85; Mismatches 154; Indels 5; Gaps 3;
QY 9 PIKPLNGLEYEQTGLFINNKFVPSKQKTFEVIINPSTEEETCHIEGREDDVEAVQAA 68
Db 59 PISPPVQVNTQ---LLIDGKFDVSAGSKTFTPLDPTGELIAHVAEHAEDINKAVHAA 115
QY 69 DRAFSNKS-WNGIDPIDRGKALYLAELIEQDKVIASIFTLONGKATSSSPG-DVRLVI 127
Db 116 PFAFTESPMPPTAVERSPIELFFALGPEHRIETAALETWIKVYAAANIEVIMVAP 176
QY 126 YFSSA-SALKI-CPWDEE-R-003YFPQGLVQGTIPWNPFLIMANKIAIAFVIG 192
Db 176 LMRYYAGWAKIDGLAVVIAAGPHVQVLIHEPVGVGQIIPWNPFLIMANKIAIAFVIG 235
QY 188 TVVLKTAESTPLSALYVSKYIPQAGTPPGVINTVSYRPGKIVVEATINHPKIKFVAFTGS 247
Db 236 LVWLKTAESTPLSALFASKLLEAGIIPGVVNVVSGPGLACAAALASHMIVKIAFVIG 295
QY 248 ATGSPHYYSAA-AQPKPVITIEGKSPNIFAALAEKLVAVONIILGIYNSHEVYAG 306
Db 246 DTGKVVLELAASNLKSVILELGGKSPNIFAALAEKLVAVONIILGIYNSHEVYAG 345
QY 307 VYVESLYNELEPTVEKAKESIPVVGQFDEHETPAGLSQMLNKLKYYVDIGKNEDAL 366
Db 356 LEVHERLYDEEVEKAKALAKAVVQDEPKNVEQVQGLDDEQENKLEKRYIKAVMSGANL 415
QY 367 TLTGGERLGSKGYFTKPTVFGGVDFMDPTVKKFTFPGVTVTKFSADEVIMNANDSEY 426
Db 416 VTGQPLRGQYGTQPTFESHVQNNPFAQEPFGVQSDITPTNDFVTPKANKAGVGI 476
QY 427 AAGIHTSNTALKAVDRVNACTVWINTYDNFHHAVPFGGFGNAGLGKREMSVDALQNY 486

```




GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:42:21, Search time: 0.2749 seconds

(without alignments)
1696 x33 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAAATCAVAPAPNQLPEVQ...YSSLAIVETKIVIKVQKKN 500

Scoring table: WUSUM62

Gap: 10 0, Gapext 0 5

Searched 908470 sws, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneset_101002.1
1: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2625	100.0	500	22	AA1974924
2	1929	73.5	517	23	AA1974916
3	1917	73.0	517	23	AA1974917
4	1866	71.1	529	24	AA1974910
5	1771	67.5	412	21	AA1974938
6	1741	66.3	521	15	AA1974972
7	1718	65.4	512	23	AA1974982
8	1718	65.4	512	23	AA1974984
9	1718	65.4	529	22	AA1974977
10	1543	58.8	538	21	AA1974912

11	1473	56.1	534	21	AA1974914
12	1470	56.0	534	22	AA1974916
13	1418.5	54.0	497	7	AA1974916
14	1305.5	49.7	446	16	AA1974914
15	1280	48.8	501	21	AA1974913
16	1280	48.8	501	21	AA1974913
17	1269	48.3	495	16	AA1974913
18	1235.5	45.7	425	22	AA1974913
19	1214	46.2	314	21	AA1974913
20	1160.5	44.2	523	23	AA1974913
21	1124.5	42.8	424	21	AA1974913
22	1098.5	41.8	909	22	AA1974913
23	1032	38.9	357	21	AA1974913
24	978.5	47.3	490	22	AA1974913
25	975.5	37.2	496	22	AA1974913
26	968.5	36.4	430	22	AA1974913
27	962.5	36.7	532	21	AA1974913
28	961.5	36.6	503	21	AA1974913
29	961.5	36.6	503	21	AA1974913
30	957.5	36.5	501	21	AA1974913
31	953	36.3	512	22	AA1974913
32	950	36.2	497	21	AA1974913
33	947.5	36.1	496	22	AA1974913
34	947.5	36.1	496	22	AA1974913
35	947	36.1	506	22	AA1974913
36	942.5	35.9	477	22	AA1974913
37	942	35.9	507	22	AA1974913
38	942	35.9	507	22	AA1974913
39	939	35.8	496	22	AA1974913
40	939	35.8	506	23	AA1974913
41	939	35.8	507	23	AA1974913
42	933.5	35.6	517	23	AA1974913
43	930	35.4	508	21	AA1974913
44	930	35.4	508	23	AA1974913
45	925	35.3	485	23	AA1974913

ALIGNMENTS

RESULT 1

AA1974924

ID: AA1974924 standard; Protein: 500 AA.

AC: AA1974924:

DE: 25-JUN-2001 (first entry)

DE: Human aldehyde dehydrogenase ALDH2 protein sequence SEQ ID NO:4.

DE: Aldehyde dehydrogenase, glycerol dehydrogenase, 3-HP, glycerol;

DE: feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;

DE: bacterial host; absorbable prosthetic device; surgical suture;

DE: beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;

DE: polyhydroxyalkonate; copolymer; lactic acid.

OS: Homo sapiens.

PN: WO200116346-A1.

PD: 08-MAR-2001.

PF: 30-AUG-2000; 2000UO-US23878.

PF: 30-AUG-1999, 99US-0151440.

PA: (WISC) WISCONSIN ALUMNI RESEARCH FOUND.

PI: Suthers PF, Cameron DC;

PI: WPI; 2001-315988/33.

DE: N-PSDB; AA1974924.

XX

PT 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 XX enzymes in the presence of glycerol or glucose -
 PS Claim 5; Page 32-35; 63pp: English.

XX The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol, (i)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (i); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (i); or (iii) carries a genetic construct
 CC which expresses the dhap gene from Klebsiella pneumoniae and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (i). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (i). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the human aldehyde dehydrogenase ALDH2, which
 CC is used in the exemplification of the present invention.

XX Sequence 500 AA;

Query Match 100.0%; Score 2625; DB 22; Length 500;
 Best Local Similarity 100.0%; Pred. No. 6e-239;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPQPEVFCNQIFINNEWHDAVSRRKFTPTVNPSTGEVLCUVAEGSKEDV 60
 DB 1 MSAAATQAVPAPQPEVFCNQIFINNEWHDAVSRRKFTPTVNPSTGEVLCUVAEGSKEDV 60
 QY 61 DKAREGPAGFQLGSPWRMDASHSGRLNRLNLIADLIERDRTYLAALLETLDNKPVIISYL 120
 DB 61 DKAREGPAGFQLGSPWRMDASHSGRLNRLNLIADLIERDRTYLAALLETLDNKPVIISYL 120
 QY 121 VLDMLVKCLRYVAGWADKYHCKTPTDGGFFSYTPHPVGVTCGLIPWNPFLLMQAWKL 180
 DB 121 VLDMLVKCLRYVAGWADKYHCKTPTDGGFFSYTPHPVGVTCGLIPWNPFLLMQAWKL 180
 QY 181 GPALATGNVVMKVAEQTPITALLYANLLKEAGFPGVVNIIVPGEGTAGAALASHESVD 240
 DB 181 GPALATGNVVMKVAEQTPITALLYANLLKEAGFPGVVNIIVPGEGTAGAALASHESVD 240
 QY 241 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLEGGKSPNIMSDADMWAVEQAHFALFNQGS 300
 DB 241 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLEGGKSPNIMSDADMWAVEQAHFALFNQGS 300
 QY 301 QCCAGSKRTVQEDUYDEFEVRSVAKAKSVGNPFUSKIFQSGPVUDETQFKKILGYINT 360
 DB 301 QCCAGSKRTVQEDUYDEFEVRSVAKAKSVGNPFUSKIFQSGPVUDETQFKKILGYINT 360
 QY 361 GKQEGAKLGGGGIADRGYFIQPTVGVGVQMGMTIAKEETFGPVNQILKEKTEIEEVVGR 420
 DB 361 GKQEGAKLGGGGIADRGYFIQPTVGVGVQMGMTIAKEETFGPVNQILKEKTEIEEVVGR 420
 QY 421 ANNSYGLAAAVFTKHLKANYLSAIGAGTWWNVYIVFGAGSPGPKYKSGSGPGLPG 480
 DB 421 ANNSTYGLAAAVFTKHLKANYLSAIGAGTWWNVYIVFGAGSPGPKYKSGSGPGLPG 480
 QY 481 YGLQAYTEVKTVTKVPQKN 500
 DB 481 YGLQAYTEVKTVTKVPQKN 500

RESULT 2

AA049516

ID AA049516 standard; Protein: 517 AA.

XX

AC AA049516;

XX 13-MAY-2002 (first entry)

XX human ALDH5 protein.

XX ALDH5; human; polymorphism; haplotype; aldehyde dehydrogenase 5;

XX binding affinity; drug targeting; alcoholism; alcohol induced disorder;
 XX antialcoholic.

XX Homo sapiens.

XX W0200192279-A2.

XX 06-DEC-2001.

XX 20-MAY-2001; 2001WO-0817253.

XX 26-MAY-2000; 2000038-20750HP.

XX (GENA-) GENA155ANT1E FIARM (N°).

XX Buda A, Finkel K, Kazemi A, Messer C, Sanchez A;

XX W01; 2002-122054/16.

XX N-PSDH; ARA99282, ARA99283.

XX New genetic variants with polymorphisms in the aldehyde dehydrogenase 5

XX (ALDH5) gene, useful for studying the function of ALDH5, and for
 XX expressing ALDH5 protein which is useful in screening drugs for
 XX treating ALDH5 related diseases -

XX Claim 40; Page 72-74; 96pp: English.

XX This invention describes a novel isolated genes and haplotypes of the

XX human aldehyde dehydrogenase 5 (ALDH5) gene containing polymorphic sites.

XX The polymorphic ALDH5 variant is useful in studying the effect of the

XX variation on the biological activity of ALDH5 and on the binding

XX affinity of candidate drugs targeting ALDH5 for the treatment of

XX alcoholism and alcohol-induced disorders. Polynucleotides comprising a

XX polymorphic gene variant or fragment may be used for therapeutic

XX purposes. ALDH5 protein isoforms may be used in assays to measure the

XX binding affinities of one or more candidate drugs targeting the ALDH5

XX protein. ALDH5 proteins may be used to generate antibodies. Haplotyping

XX method can be used by scientists to validate ALDH5 as a candidate

XX target for treating a specific condition or disease predicted to be

XX associated with ALDH5 activity, and in the design of clinical trials of

XX candidate drugs for treating a specific condition or disease predicted

XX to be associated with ALDH5 activity. Information on polymorphisms on

XX the ALDH5 gene can be applied for studying the biological function of

XX ALDH5 as well as in identifying drugs targeting this protein for the

XX treatment of disorders related to its abnormal expression or function.

XX The products of the invention have an allelic activity. This sequence

XX represents the human ALDH5 protein described in the disclosure of the

XX invention.

XX Sequence 517 AA;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

182	QY	PALATGNVVVVKVAEQPTLIALYVANLKEAGFPFGVWNI	VFGPFTAGAAZASHEDYDK	241
198	Db	PALATGNTVVKVAEQPTLSALYLASLKEAGFPFGVWNI	ITGYGPTAGAAQAQHMVDK	257
242	QY	VAFGTSTETGRTVAAFGSNLKRVTLELGSXFNILMS	LALUMUWAVEVAIFALFN34	301
258	Db	VAFGTSTEVGHILTKAAGSNLKRVTIEGGRSPSLV	LALAMEHAQECHLEAFNN3Q	317
302	QY	CCAGSRTFVGPDYIDFVVKSVAKAKSKVGVNPFDSK	TEGQVUETQFKKL3AY	361
318	Db	CCAGSRTFVPESTYNEFLERTVEKAKQKVGNPFELL	TQGSQVLUKLFERVL3YIUJ3	377
362	QY	KQEGAKLLCGGSIANDRGYFIQVTFSTFVQJGNTIA	KKEIFGVWQIUKFKRTIEVW74A	421
378	Db	KQEGAKLLCGSERFERGEGFTIKTPTFGVQDDMRI	AKKEIFGPVQVPLFKFKKIEBW5KA	437
422	QY	NNSTVGLAAAVFTKDLLDKANVLSQALCAGTVWVNC	YDVFAGSPGSGYKMSGSGRELGEY	481
438	Db	NNRTVGLAAAVFTKDLLDKANVFTQALCAGTVWVNT	YNIVTCHTPPGGPKESONGRELGED	497
482	QY	GLQAYTEVKTVTVKVPQKN	500	
498	Db	GLKAYTEVKTVTKVPQKN	516	

RESULT 3	
AAM49517	
ID	AAM49517 standard; Protein; 517 AA.
XX	
AC	AAM49517;
XX	
DT	13-MAY-2002 (first entry)
XX	
DE	Human ALDH5 protein polymorphic variant.

PT	treating ALDH5-related diseases -
XX	Enucleated; Fig 3; 6pp; English
FX	
PS	This invention describes a novel isolated genes and haplotypes of the
CC	human aldehyde dehydrogenase " (ALDH5) gene containing polymorphic sites,
CCC	the polymorphic ALDH5 variant is useful in studying the effect of the
CCC	variation on the biological activity of ALDH5 and on the binding
CCC	affinity of candidate drugs targeting ALDH5 for the treatment of
CCC	alcoholism and alcohol-induced disorders. Polynucleotides comprising a
CCC	polymorphic gene variant or fragment may be used for therapeutic
CCC	purposes. ALDH5 protein isoforms may be used in assays to measure the
CCC	binding affinities of one or more candidate drugs targeting the ALDH5
CCC	protein. ALDH5 proteins may be used to generate antibodies. Haplotyping
CCC	method can be used by scientists to validate ALDH5 as a candidate
CCC	target for treating a specific condition or disease predicted to be
CCC	associated with ALDH5 activity, and in the design of clinical trials of
CCC	candidate drugs for treating a specific condition or disease predicted
CCC	to be associated with ALDH5 activity. Information on polymorphisms on
CCC	the ALDH5 gene can be applied for studying the biological function of
CCC	ALDH5 as well as in identifying drugs targeting this protein for the
CCC	treatment of disorders related to its abnormal expression or function.
CCC	The products of the invention have anti-alcoholic activity. This sequence
CCC	represents the human ALDH5 polymorphic variant protein described in the
CCC	disclosure of the invention.
XX	Sequence 517 AA;
SQ	

RESULT 4
ABB60140
14 ABB60140 standard, Proteins, 520 AA
XX


```

Db      19 ALPRIRNLVKKFTKFINNEWHESKGFATCNSTRPQICEVEEGKPVVUKAVEAA 78
QY      68 PGAFQLGSPWRMDASGRLLNPIADLIERPTYIAAETLNGSKPYVISYLVLEDMVL 127
Db      79 QVAFQSGSPWRRLDASGRLLHQDLVDRDRATLAALETMDTGKPFLLHAFIDLEGCI 138
QY      128 KCLRYAGWADKYHGKTIPIIDGDFFSYTRHEPVGVSQIIPWNPFLMLQAWKLGPAATG 187
Db      139 RFLRYFAGWADKIQKTIPIIDWVCFTRHEPVGVSQIIPWNPFLMLQAWKLGPAATG 198
QY      188 NVVMKVAEQPTLITALLYANLKEAGPPGVVNIIVFGFGTAGAATASHEDYKVAFTGS 247
Db      199 NTMVLKDAEQPTLITALLYANLKEAGPPGVVNIIVFGFGTAGAATASHEDYKVAFTGS 258
QY      248 TEIGRVITQVAGSSNIPKVTIETGCKSPNIIIMSDALMDWAVEGAHFALEFNGQCCAGS 307
Db      259 TEVSKLVKEAASRNIPKVTIETGCKSPNIIIMSDALMDWAVEGAHFALEFNGQCCAGS 318
QY      308 FTEVQERTYDFEVVPSVAPAKSPVVGNPFSKTEQGVVETGFKKILYINFRKQKSAK 367
Db      319 RVFVEEQVYSFVKRSVEYAKKRPVGVLPVUKTEGQPIIQKUFUKLLEIESGKKEGAK 378
QY      368 LIAFGSTAAEPSEYELQTVRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 427
Db      379 LECGGSAMENKGLFKPTVSEVTDNNMPIAKREFTPGVQVPIIKFKSIEEVIKPANSTDVG 438
QY      428 LAAAVFTKLDKALKLASALESSTVWVNYGVDPVCAQSPDQVYKMSISGPELGYGLAYT 487
Db      439 LTAAVFTKLDKALKLASALESSTVWVNYGVDPVCAQSPDQVYKMSISGPELGYGLAYT 498
QY      488 EVKTVTVKVPQKN 500
Db      499 EVKTVTVKVPQKN 511

```

RESULT 9

ABG06577
ID ABG06577 standard; protein: 529 AA.

AC ABG06577;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #6568.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement, medical imaging, diagnostic, genetic disorder

OS Homo sapiens.

PN WO2001/5067 A2.

PO 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0649167.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DF WPI; 2001-639362/73.

DR N-PSDB; AAS0764.

PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity

PS claim 20; SEQ ID NO 364a-b; language: English

XX the invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity
CC the polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. Amino acid sequences of the invention
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SEQ Sequence 529 AA;

Query Match 65.4% Score 1718; DB 22; Length 529;
Best Local Similarity 65.5% Pred. No. 3 le-153;
Matches 22; Conservative 66; Mismatches 104; Indels 0; Gaps 0;

```

QY      1 AVPAWNPQPEVFNQIPIINNEWHIAVSPKTIPIVNFSTGEVILVAEGSKHIVLRAKPPSP 67
Db      36 ALPRIRNLVKKFTKFINNEWHESKGFATCNSTRPQICEVEEGKPVVUKAVEAA 95
QY      68 PGAFQLGSPWRMDASGRLLNPIADLIERPTYIAAETLNGSKPYVISYLVLEDMVL 127
Db      96 QVAFQSGSPWRRLDASGRLLHQDLVDRDRATLAALETMDTGKPFLLHAFIDLEGCI 155
QY      128 KCLRYAGWADKYHGKTIPIIDGDFFSYTRHEPVGVSQIIPWNPFLMLQAWKLGPAATG 187
Db      156 PTLRYFAGWADKIQKTIPIIDWVCFTRHEPVGVSQIIPWNPFLMLQAWKLGPAATG 215
QY      188 NVVMKVAEQPTLITALLYANLKEAGPPGVVNIIVFGFGTAGAATASHEDYKVAFTGS 247
Db      216 NTMVLKDAEQPTLITALLYANLKEAGPPGVVNIIVFGFGTAGAATASHEDYKVAFTGS 275
QY      248 TEIGRVITQVAGSSNIPKVTIETGCKSPNIIIMSDALMDWAVEGAHFALEFNGQCCAGS 307
Db      276 TEVSKLVKEAASRNIPKVTIETGCKSPNIIIMSDALMDWAVEGAHFALEFNGQCCAGS 335
QY      308 FTEVQERTYDFEVVPSVAPAKSPVVGNPFSKTEQGVVETGFKKILYINFRKQKSAK 367
Db      336 RVFVEEQVYSFVKRSVEYAKKRPVGVLPVUKTEGQPIIQKUFUKLLEIESGKKEGAK 395
QY      368 LIAFGSTAAEPSEYELQTVRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 427
Db      396 LECGGSAMENKGLFKPTVSEVTDNNMPIAKREFTPGVQVPIIKFKSIEEVIKPANSTDVG 455
QY      428 LAAAVFTKLDKALKLASALESSTVWVNYGVDPVCAQSPDQVYKMSISGPELGYGLAYT 487
Db      456 LTAAVFTKLDKALKLASALESSTVWVNYGVDPVCAQSPDQVYKMSISGPELGYGLAYT 515
QY      488 EVKTVTVKVPQKN 500
Db      516 EVKTVTVKVPQKN 528

```

RESULT 10

AAV67412
ID AAV67412 standard; protein: 538 AA.

AC AAV67412;

DT 12-MAY-2000 (first entry)


```

Db      197  LVAAVLKAGPPGVINVSQFVACAAALSSHPVTKVAFITGVGTILKAAASSN 256
QY      263  LKRVILEGKSPINILMSIALMDWAVEQAHPALFNNQCCAGSRFEVVEDIYDFVVR 322
Db      257  LKKVILEGKSPNIVFETADIDNAISWVNFIFENHGQCCAGSRVYVQESIYDFVQK 316
QY      323  SVARAKSRVVSAPPUSKTEQSPQVDETQFKKILCY INQKQESAKLLGQSTIAADPGYFI 382
Db      317  FFRKAKNVVGVPPFAADIFQAPQSVKQGFQFMKYLGAKLQATAIVETGSRKGGKSYFI 476
QY      383  OPTVEGDVGDGMTIAKRFIFGPMQILKFKTIEEVSVPANNSTYGLAAAVETKDLQKANY 442
Db      377  EPTILSNVTEDMKIVKEEIFGPMVCSIAKFKTKEDAIKLGNASIVGLAAAVHTRNLNTALE 436
QY      443  LSAALGALIVWVNYVDFCAVSPSPGSKMSISGFFLQYGLLQAYTEVKTIVTVAV 496
Db      437  VSNLAKAGTVWVNTYTLHHQMPGGYKESGIGBELGADALANTYTKTVSTIRI 490

RESULT 15
AAG36239
ID   AAG36239 standard, Protein, 501 AA.
XX   AC   AAG36239;
XX   DT   18-OCT-2000 (first entry)
XX   DF   Arabidopsis thaliana protein fragment SEQ ID NO: 44363.
XX   KW   Protein id-confirmation; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX   OS   Arabidopsis thaliana.
XX   PN   EP1033405-A2.
XX   PD   06-SEP-2000
XX   PF   25-FEB-2000; 2000FF-0301430
XX   PP   25-FEB-1999; 99US-0121825
PP   05-MAP-1999; 99US-0123180
PP   09-MAP-1999; 99US-0123548
PP   23-MAP-1999; 99US-0125788
PP   25-MAP-1999; 99US-0126264
PP   29-MAP-1999; 99US-0126785
PP   01-APR-1999; 99US-0127462
PP   06-APR-1999; 99US-0128234
PP   08-APR-1999; 99US-0128714
PP   16-APR-1999; 99US-0129845
PP   19-APR-1999; 99US-0130077
PP   21-APR-1999; 99US-0130444
PP   23-APR-1999; 99US-0130510
PP   23-APR-1999; 99US-0130801
PP   30-APR-1999; 99US-0131449
PP   30-APR-1999; 99US-0132048
PP   04-MAY-1999; 99US-0132407
PP   04-MAY-1999; 99US-0132484
PP   05-MAY-1999; 99US-0132485
PP   06-MAY-1999; 99US-0132486
PP   06-MAY-1999; 99US-0132487
PP   07-MAY-1999; 99US-0132863
PP   11-MAY-1999; 99US-0132956
PP   14-MAY-1999; 99US-0134218
PP   14-MAY-1999; 99US-0134219
PP   14-MAY-1999; 99US-0134221
PP   14-MAY-1999; 99US-0134370
PP   18-MAY-1999; 99US-0134768
PP   19-MAY-1999; 99US-0134941
PP   20-MAY-1999; 99US-0135124
PP   21-MAY-1999; 99US-0135353
PP   24-MAY-1999; 99US-0135624
PP   26-MAY-1999; 99US-0136021
PP   27-MAY-1999; 99US-0136392
PP   28-MAY-1999; 99US-0136782
PP   01-JUN-1999; 99US-0137222
PP   03-JUN-1999; 99US-0137528
PP   04-JUN-1999; 99US-0137502
PP   07-JUN-1999; 99US-0137724
PP   08-JUN-1999; 99US-0138094
PP   10-JUN-1999; 99US-0138540
PP   10-JUN-1999; 99US-0138847
PP   14-JUN-1999; 99US-0139119
PP   16-JUN-1999; 99US-0139452
PP   16-JUN-1999; 99US-0139453
PP   17-JUN-1999; 99US-0139492
PP   18-JUN-1999; 99US-0139454
PP   18-JUN-1999; 99US-0139455
PP   18-JUN-1999; 99US-0139456
PP   18-JUN-1999; 99US-0139457
PP   18-JUN-1999; 99US-0139458
PP   18-JUN-1999; 99US-0139459
PP   18-JUN-1999; 99US-0139460
PP   18-JUN-1999; 99US-0139461
PP   18-JUN-1999; 99US-0139462
PP   18-JUN-1999; 99US-0139463
PP   18-JUN-1999; 99US-0139750
PP   18-JUN-1999; 99US-0139763
PP   21-JUN-1999; 99US-0139817
PP   22-JUN-1999; 99US-0139899
PP   23-JUN-1999; 99US-0140353
PP   24-JUN-1999; 99US-0140354
PP   24-JUN-1999; 99US-0140695
PP   28-JUN-1999; 99US-0140823
PP   29-JUN-1999; 99US-0140991
PP   30-JUN-1999; 99US-0141287
PP   01-JUL-1999; 99US-0141842
PP   02-JUL-1999; 99US-0142154
PP   06-JUL-1999; 99US-0142055
PP   06-JUL-1999; 99US-0142390
PP   08-JUL-1999; 99US-0142803
PP   09-JUL-1999; 99US-0142920
PP   12-JUL-1999; 99US-0142977
PP   13-JUL-1999; 99US-0143542
PP   14-JUL-1999; 99US-0143624
PP   15-JUL-1999; 99US-0144005
PP   16-JUL-1999; 99US-0144085
PP   16-JUL-1999; 99US-0144086
PP   19-JUL-1999; 99US-0144325
PP   19-JUL-1999; 99US-0144331
PP   19-JUL-1999; 99US-0144332
PP   19-JUL-1999; 99US-0144334
PP   19-JUL-1999; 99US-0144334
PP   19-JUL-1999; 99US-0144335
PP   20-JUL-1999; 99US-0144352
PP   20-JUL-1999; 99US-0144542
PP   20-JUL-1999; 99US-0144884
PP   21-JUL-1999; 99US-0144814
PP   21-JUL-1999; 99US-0145086
PP   21-JUL-1999; 99US-0145088
PP   22-JUL-1999; 99US-0145085
PP   22-JUL-1999; 99US-0145087
PP   22-JUL-1999; 99US-0145089
PP   22-JUL-1999; 99US-0145192
PP   23-JUL-1999; 99US-0145145
PP   23-JUL-1999; 99US-0145218
PP   23-JUL-1999; 99US-0145224
PP   26-JUL-1999; 99US-0145276
PP   27-JUL-1999; 99US-0145913
PP   27-JUL-1999; 99US-0145918
PP   27-JUL-1999; 99US-0145919
PP   28-JUL-1999; 99US-0145951
PP   02-AUG-1999; 99US-0146386
PP   02-AUG-1999; 99US-0146486

```



```

Db 77 DKAVKAAQAFQIGSPWRTMDASGRLLYLKLADLIERORLLATMESMNGCKLYSNA 136
QY 119 YLVOLDVWIKLRYAGWACKY--HCKTIPITGDFESYTHPEPVWCGSILIPWNEPLLMQ 176
Db 137 YLNULAARIKILRYAGWACKY--HCKTIPITGDFESYTHPEPVWCGSILIPWNEPLLMQ 195
QY 177 AWKLGALATGNVVMKVAEDELITALYVANLIKKEAGPPGVVNVIVPGPGTAGAAIASH 236
Db 197 IWKIGPALSNGNVVVKPAQOTPLTALHVASLIKKEAGPPGVVNVIVPGPGTAGAAIASH 256
QY 237 EDVQKVAFTSIEIGRVIOVAAAGSNLKVITLGLGKSPNLIIMSDADMWAVEQAHFALF 296
Db 257 MDIDKVAFTGSTVGKLIKEAAGSNLKVITLGLGKSPCIVLADADLONAVEPAHGVF 316
QY 297 FNOQCQACASRTFVQETITFVFPVSVAKAKSPVWVNFPSKFPQVQVFTQKPKHLS 356
Db 317 YHQGCCIAASRTFVEESYDEVRBSVERAKKYYIIGNPITPQVTOGPQIQKQYDKIL 376
QY 357 YINTGKQETAKLQAGSIAALQGYFTQPTVEGQWQGMIAKEELTGPVQMIKFKITEE 416
Db 377 LIESKKEGAKLECGSPWCKNGYFVQPTVFSNVTDENMIKKEELTGPVQMIKFKSLDD 436
QY 417 VVGRANNSTVGLAAVFTKDLDRANKYLSQALQAGTVWVNCYDVFQAQSPFGQYKMSGGR 476
Db 437 VIKRANNTFYGLSAGVETKIDKRAITISSALQAGTVWVNCYGSVGNQCFPGGKMSGNRP 496
QY 477 ELGFGYGLQAYTEVKTIVTKVPOKN 500
Db 497 ELGEGHPHEVTEVKTIVTKVISQKN 520

```

RESULT 2

US-09-134-001C-4383

; Sequence 4383, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4383

; LENGTH: 506

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4383

Query Match 35.8%; Score 949; DB 4; Length 506;

Best Local Similarity 41.8%; Pred No 1,5e-91;

Matches 207; Conservative 7%; Mismatches 186; Indels 10; Gaps 6;

```

QY 21 NQIFINNEHDAVSRKTEFTNPFGEVICOVAEGKRTIVIKAREPPEAFGLSPWPKRM 80
Db 19 NQVIDGFWVSSNKNTRDIINPYNGETITFVAEGTKEDVERAIIAARRSEFGE-WSLE 77

```

```

QY 81 DASHSRLINKLAULERERTYIALETGNKRPVIVSIYGLIMVLCIERYAGWAKY 140
Db 78 TSEVHKKVRVAADKIKFNPEFAKIPITDCKTIEFSY-ADMDIHNHVPYFACLAIDK 135

```

```

QY 141 HGKTI--PIDGDFESYTRIEPVGVCQQLIPWNEPLLMQAWKLGALATGNVVMKVAFO 198
Db 137 GGEIINSPIP-NAESKVVKEPVQVVOITPWNYPILQASWKIAPALATCCSILVWKPSEIT 195

```

```

QY 199 PITAIYVANLIKAGPPGVVNVIVPGPGTAGAAIASHEDVIVKVAFTSTEGTGVICVAA 258
Db 196 PITTIVPPEIMFVFPKPTIINLIVICAGSEVGVWVS-HEFVIVSFTETKHKMKQA 255

```

```

QY 259 GSSNLKRVITELGKSPNLIIMSDADMWAVEJAHFALFENQVQVCAASKRTVQVDFAYDE 318
Db 256 -ANBETIWALEGGERNINTEFTAAQETAAVPAALNAGYTHAGQVSAKSPETIVNCDKOR 314
QY 319 EVFESVARAKSPVWVNFPSKFPQVQVFTQKPKHLSINLQKGGZAKLLAGAGGIA 376
Db 317 YHQGCCIAASRTFVEESYDEVRBSVERAKKYYIIGNPITPQVTOGPQIQKQYDKIL 376
QY 376 -AUGGYFTQVTPGVODGMIKAKEIFPQVWQMIKFKRTIEVVVGRANNSTVGLAAVET 444
Db 375 DLQAGLFFETVITDCTSMKRVQERFQVGVVTFEFAEEAATKLANUSTVGLAAVET 444
QY 435 KILUKANYLSQALQAGTVWVNCYDVFQAQSPFGQYKMSGGRFELGFGYGLQAYTEVKT 492
Db 437 YHQGCCIAASRTFVEESYDEVRBSVERAKKYYIIGNPITPQVTOGPQIQKQYDKIL 492

```

RESULT 3

US-09-655-270A-9

; Sequence 9, Application US/09655270A

; Patent No. 6329151

; GENERAL INFORMATION:

; APPLICANT: Rouviere, Pierre E.

; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Proteins

; FILE REFERENCE: R01011 US NA

; CURRENT APPLICATION NUMBER: US/09/655,270A

; PRIOR FILING DATE: 2000-09-05

; PRIOR FILING DATE: 60/120,702

; PRIOR FILING DATE: 1999-February 19

; PRIOR FILING DATE: 60/152,542

; PRIOR FILING DATE: 1999-September 04

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: MicroSoft Office 97

; SEQ ID NO 9

; LENGTH: 508

; TYPE: PRT

; ORGANISM: Rhodococcus erythropolis HL PM-1

US-09-655-270A-9

Query Match 35.4%; Score 930; DB 4; Length 508;

Best Local Similarity 37.9%; Pred No 1,6e-90;

Matches 194; Conservative 92; Mismatches 200; Indels 26; Gaps 4;

```

QY 9 VPAPNOQPEVFNQ-----V-----VFNNHWDVSRKTEFTVNPSPGEVIG 50
Db 1 LPTPSSPPHTCTGCGAEEKTRIVQALTISSVPLVICHQITLPSGLATLDSINPAKSHLA 60

```

```

QY 51 QVAPGTFHVDVAPPDPPGAPALGSSWPPPMASPSPTINPQATFDPQPTVLAALETLD 110
Db 61 SVAETAAIIVARAVEAKAA--ARTQWRMRVQORTLMPRYAALFEERKTELADQDSRO 117

```

```

QY 111 NCKPVVISVIMQDMVKRLRYVYAGWCKYKRIITDQDFSVTRHFWVQVQITLWN 170
Db 118 MKQPTRESLGIDLPIMETLETFAGLVTKTEGRTTPAIRSRLINYLREPTCAVVAITLWN 177

```

```

QY 171 FTLLMQAWKLGALATGNVVMKVAQDPLITALYVANLIKKEAGTGVVNVIVPGPGTAG 240
Db 178 FPAVAVVWKAVALAMGNALVLRPAQALVIVVAGLALAGLFTPLVNVLPKGSVAG 247

```

```

QY 231 AAIAASHEDVIVKVAFTSTEGTGVICVAAAGSSNLIKVTITELGKSPNLIIMSDADMWAVEQ 290
Db 236 NALVQHSWCKVIFGSEVYVQCGGMA-ADPLTIASLEDEGSAKAVAGRSKRAVAV 296

```

```

QY 291 AHFALFPNQCQVCAASKRTFVQDITDFEVVHVSVAKAKSPVWVNFPSKFPQVQVFTQ 350
Db 297 VGUAMYSNOGETCTAHSRLVVERPLYDEKVELVQARVAAKVGFDLPDITETGLISARQ 356

```

```

QY 451 FKKILCYINTGQEGAKLIGAGI-----AAGTCFETQTVFDVQDGMTAKKEIFSPVM 406
Db 457 FESVHSVYVSGTFEGATISGQAGSPGAPGAYRPIETESVIAUMPAARELPQVPL 416

```

```

QY 497 QILKKEKLEEVGVANNSTVGLAAVETKILQVANYLSQALQATVWVNVIVPGPGTAG 492

```


QY 84 HSGRLNPLADIFEDRTYLAALETHINCKPYYVLSYLVKLLMWLKLPLYVYAGWAKYHCK 143
 DB 76 DPAAVILKAAGLLPERRDDIAYWVINGKP-ISOAKGEIDHCHACFEMAAGARMMLHGD 134
 QY 144 TIPIDGG-FFSYTHPEPVGCGQITIPWNPFI LKQAWKFLPALATGNVWVKNVAAQFELIA 202
 DB 135 TFNNLGEGLFGMWLRPIGVVGLITPWNFPFIMILCERAPFIASGCTLVVKKPAFVTSATT 194
 QY 203 LYVANLKEAGFPFGVNVIVGPGPTAGAAALASHEDYDKVAFGTGSGEIGRVIOVAGSSN 262
 DB 195 LLLAEILADAGLPGKGVNVTGTGRVGOAMIEHQDIDMLSTGTCGVGKSCILHAADS 254
 QY 263 LKPVITRIGKSPNTKMSDAMTWAVEQAHLALFFKGSJAGSEIFVLEEDYDEVVE 322
 DB 255 LKKLGLGKGNPIVWFADSNLEDAADAVAFISFNTGCGCVSSSKLIVERSVAEKEFL 314
 QY 323 SVARAKSVGNPNDSKTEQGPQVDETFQFKKILGYINTGKOGAKLLCGGGIA-ADRGYF 381
 DB 315 VPKMEKIRVGPDEPDTQIGAITTEAQNRIILDYIAKGAEGAKLLCGGIVDFGKGY 374
 QY 382 IOPTVFGVODGDMTIAKEEIPGVMQILKTKTIEEVVGRANNSTYGLAAVFTKDLDRAN 441
 DB 375 IOPTLFTDVKPSMGIADEIPGVLASPHFDTVDEAIAIANDTVYGLAASVWSKIDKAL 434
 QY 442 YLSQALQAGTIVWVNYVFRQAQSPFGKYSKSGSPRELGEYSLQAYTEVKITIVKVPQKN 500
 DB 435 AVTRVRAGRWVNTINSGGPELGGFKQSGWKGAGLYGVEXYQIKSVHJETGKRS 493

RESULT 8

US-08-942-673-2

Sequence 2, Application US/08942673

Patent No. 5851292

GENERAL INFORMATION

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshinori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: L-sorbose body hydrogenase and No. 5851292-1

TITLE OF INVENTION: L-sorbose dehydrogenase obtained from Glucanobacter

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/942,673

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/513,841

FILING DATE: 01-NOV-1995

APPLICATION NUMBER: 08/930,700 q

FILING DATE: 08-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 241871/1993

FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-q00-0 pct

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-4240
 TELEX: 248855 OPAT OR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 ORGANISM: Glucanobacter oxydans
 STRAIN: T-100
 FEATURE:
 NAME/KEY: cat peptide
 LOCATION: 1-497
 IDENTIFICATION METHOD: experimentally
 US-08-942-673-2

Query Match 34.4% Score 932; DB 2; Length 497;

Best Local Similarity 39.5%; Pred. No. 1, 3e-87;

Matches 189; Conservative 98; Mismatches 186; Indels 6; Gaps 5;

QY 24 PINNWHVANSPKTPTPTVNSTEVIQCVAFQKKEIVKKAFFSPGAFQLGSPWRPMDS 83
 DB 19 FIDRWP--AKGKPFDESSPAHEVPVTRIP--TEIDLEAFAAARAFENGSG-WAGLAA 75
 QY 84 HSCPLNPLADIFEDRTYLAALFTINGKPPYVLSYLVKLLMWLKLPLYVYAGWAKYHCK 143
 DB 76 DPAAVILKAAGLLPERRDDIAYWVINGKP-ISOAKGEIDHCHACFEMAAGARMMLHGD 134
 QY 144 TIPIDGG-FFSYTHPEPVGCGQITIPWNPFI LKQAWKFLPALATGNVWVKNVAAQFELIA 202
 DB 135 TFNNLGEGLFGMWLRPIGVVGLITPWNFPFIMILCERAPFIASGCTLVVKKPAFVTSATT 194
 QY 203 LYVANLKEAGFPFGVNVIVGPGPTAGAAALASHEDYDKVAFGTGSGEIGRVIOVAGSSN 262
 DB 195 LLLAEILADAGLPGKGVNVTGTGRVGOAMIEHQDIDMLSTGTCGVGKSCILHAADS 254
 QY 263 LKPVITRIGKSPNTKMSDAMTWAVEQAHLALFFKGSJAGSEIFVLEEDYDEVVE 322
 DB 255 LKKLGLGKGNPIVWFADSNLEDAADAVAFISFNTGCGCVSSSKLIVERSVAEKEFL 314
 QY 323 SVARAKSVGNPNDSKTEQGPQVDETFQFKKILGYINTGKOGAKLLCGGGIA-ADRGYF 381
 DB 315 VPKMEKIRVGPDEPDTQIGAITTEAQNRIILDYIAKGAEGAKLLCGGIVDFGKGY 374
 QY 382 IOPTVFGVODGDMTIAKEEIPGVMQILKTKTIEEVVGRANNSTYGLAAVFTKDLDRAN 441
 DB 375 IOPTLFTDVKPSMGIADEIPGVLASPHFDTVDEAIAIANDTVYGLAASVWSKIDKAL 434
 QY 442 YLSQALQAGTIVWVNYVFRQAQSPFGKYSKSGSPRELGEYSLQAYTEVKITIVKVPQKN 500
 DB 435 AVTRVRAGRWVNTINSGGPELGGFKQSGWKGAGLYGVEXYQIKSVHJETGKRS 493

RESULT 9

US-09-118-317-2

Sequence 2, Application US/09118317

Patent No. 6197562

GENERAL INFORMATION:

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshinori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: No. 6197562-1 L-sorbose dehydrogenase and No. 6197562-2

TITLE OF INVENTION: L-sorbose dehydrogenase obtained from Glucanobacter

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

```

: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1 44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: MS-DOS Editor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/118,317
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/513,841
: FILING DATE: 01-NOV-1995
: APPLICATION NUMBER: UK 9404700.9
: FILING DATE: 08-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 241851/1993
: FILING DATE: 28-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: NORMAN F. ORLON
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 18-909-0 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-413-3000
: TELEFAX: 703-413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 497 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Gluconobacter oxydans
: STRAIN: T-100
: FEATURE:
: NAME/KEY: mat peptide
: LOCATION: 1..497
: IDENTIFICATION METHOD: experimentally
US-09-118-317-2

```

```

Query Match 44.4%; Score 902; DB 4; Length 497;
Best Local Similarity 39.5%; Pred. No. 1.3e-87;
Matches 189; Conservative 98; Mismatches 186; Indels 6; Gaps 5;

QY 24 FINNEWHDAVSRKTPPTVNPSTGEVIGQVAKGDKEDKAKRGKPPAFGLSGPWRMDA 83
DB 19 FIDGHR--AGKDFDRSSPAHVVPTRIPKTPPLDEAVAAAPPAFPNPS-WAGLAAA 75
QY 84 HSGRLINRIADIETDRITYLAALETLDSKPYVTSYLVUIMVLKLPYYAGWALKYHSG 143
DB 76 DRAAVLLKAAGLLIRERDDIATWEVLENSKP-LSQAKCELDHC IACFEHMAAGAAKMLHGD 134
QY 144 TLPIDKDFPSYTRHEPVCVCGQLIPWNPFLLMQAWKLGCPALATGNNVVMKVAQTPLTA 202
DB 135 TFNNLREGFLRWVREPICVGLITPWNPPFEMILCEAPFFILASCTILVWKPAEVISAII 194
QY 203 LYVANI.IKAGPGPVGNNIVPGFGPTAGAAIASHHEWIKVAFTEGTEIGRVIQVAAAGSSN 262
DB 195 LLLAFILATAGIPKTVENVVTITGTGTGAMTEHGLIMLSFISTGVGKSCIIHAAALSN 254
QY 263 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 322
DB 255 LKKLIGHLELGNKPIVVFAUSNLEDAADAVAFGISHTGQCVCSSSRLIVERSVAEKPERL 314
QY 323 SVARAKSRVVGPNFDSKTEGQPVDETOFKKTLIGYINTGKGEGAKLLCGGSLA-ADRGYF 381
DB 315 VPKMKKIRVGDPDFDPTDITGAITTEAQNKTILDYIAKGAEGAKLLCGGGIVDFGKGQY 374
QY 382 IQPTFGVDGDMTAKERIFGVVMQILKFKTIERVVGGRANNSIVGLAAAVFTKDLDPKAN 441
DB 375 IQPTLFTDVKPSMGIAERDIFGVLASRHFEDTVDATAIANDTVYGLAASWMSKDIKAL 434

```

```

QY 442 YLSQALQASTVWVNYVEVGAQSPGGYKMSGSGHEI-GYGLQAVTEVTVVVKVQGN 500
DB 435 AVTRRVAGKFWNTIMSGPPTPLGGPKQSGWGFHAYLYGVETVQLKSVHIEGKKS 494

RESULT 10
US-09-134-001C-4541
: Sequence 4541, Application US/09/144001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/144,001C
: CURRENT FILING DATE: 1998-08-14
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4541
: LENGTH: 510
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4541

```

```

Query Match 33.7%; Score 884; DB 4; Length 510;
Best Local Similarity 38.4%; Pred. No. 1.1e-85;
Matches 184; Conservative 94; Mismatches 187; Indels 14; Gaps 6;

QY 23 IFINNEWHDAVSRKTPPTVNPSTGEVIGQVAKGDKEDKAKRGKPPAFGLSGPWRMDA 82
DB 29 LFINNEFOASUSGETLIVSNFANGEDAKAVAKGKWDKAVQAALDAF DSWSKSR 85
QY 83 SHSGRLINRIADIETDRITYLAALETLDSKPYVTSYLVUIMVLKLPYYAGWALKYHSG 142
DB 86 EFPADVLEISPPRTHPKTHIATVESLQNGKIYPTSTIIVIGAAAGPKYFASVLIIDGG 145
QY 143 KTIPIKCDFFSYTRHEPVCVCGQLIPWNPFLLMQAWKLGCPALATGNNVVMKVAQTPLTA 202
DB 146 SVNEILQKRTMS:VNNHVAWGLVVAWNIITLLASWEGPILAAARVVIQPGSGSTFEE 205
QY 203 LYVANI.IKAGPGPVGNNIVPGFGPTAGAAIASHHEWIKVAFTEGTEIGRVIQVAAAGSSN 262
DB 206 LLLAFILATAGIPKTVENVVTITGTGTGAMTEHGLIMLSFISTGVGKSCIIHAAALSN 254
QY 263 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 322
DB 264 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 322
QY 323 SVARAKSRVVGPNFDSKTEGQPVDETOFKKTLIGYINTGKGEGAKLLCGGSLA-ADRGYF 381
DB 324 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 382
QY 375 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 435
DB 383 LKRVTELGKGSINIIMSDADMWAVEQAHAFALEFNQGGVATAGASRFPVQELIYDERVVR 441
QY 442 YLSQALQASTVWVNYVEVGAQSPGGYKMSGSGHEI-GYGLQAVTEVTVVVKVQGN 500
DB 442 YLSQALQASTVWVNYVEVGAQSPGGYKMSGSGHEI-GYGLQAVTEVTVVVKVQGN 500

```

```

RESULT 11
US-09-151-224F-5
: Sequence 5, Application US/09/151224F
: Patent No. 6388171
: GENERAL INFORMATION:
: APPLICANT: Duivick, Jon
: APPLICANT: Maddox, Joyce
: APPLICANT: Gilliam, Jacob

```

; APPLICANT: Folkerts, Otto
 ; APPLICANT: Crasta, Oswald R.
 ; TITLE OF INVENTION: Compositions and Methods for Immunisation
 ; TITLE OF INVENTION: Detoxification
 ; FILE REFERENCE: 5718-111
 ; CURRENT APPLICATION NUMBER: US/09/451,224E
 ; CURRENT FILING DATE: 1999-07-12
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 487
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 ; US-09-351-224E 5

Query Match 31.2%, Score 818 5, DB 4, Length 487;
 Best Local Similarity 40.0%, Pred. No. 1.1e-78;
 Matches 195; Conservative 86; Mismatches 193; Indels 13; Gaps 11;
 QY 16 PEVFCNQIFINNEHDAVSRIKFTVNP-SIGEVICQVAGDKEDVDKAREGRGAFOLG 74
 DB 5 PRFYKSEIFINNEFVSSKSGSEPLITINPDESTVATDVHANAADVSAVASQAVKKG 64
 QY 75 SWRRMDASHGRLNRLADLIERDRTYLALETLDNGKPYVSYLVLDLDMVZLKLRYYA 134
 DB 65 -PWKFTGAQRAA'MLKPADLAENAEKLAPESTPTGPPVSMITHPDIPNMYSVRYA 123
 QY 135 GWADKYHGKTIPIIDGDFSYTRHPPVGVCGGOLIPNFPLLMQAWKLGCPALATGNVVMKV 194
 DB 124 GWADKIAKTFPEFGKKNW-PYRPMGVAGTASWNAFTFYVGMKTAAPAAAG'STIFKA 142
 QY 195 ABOTRLTALYVANLIKAEAGFPFGVGVNIVFSGPTAGAAASHEDVDKVAFTSTGTEGRVI 254
 DB 183 SEKSLPLGLAPLFAEAGFPVGVVFLTGARVI-GEALASHMDIAKISITRSVGGRAV 241
 QY 255 QVAASSNLKRVTLGKGSNI'IMSADMDWAVEQAHPALFNU-QGCCACSGKRVKQ- 312
 DB 242 KZATTKSNMKRVTLGKGR-PTIVNEAPLE-PQSGERAKDFSKFQGLWVPSGLIVQW 238
 QY 313 EDIYDEF--VPSVAPAKSPVVS NPLSKIEHQCPVLEIGKFF IIGVINTGKPPAKIL 464
 DB 299 SNAEKFDVVRWSSGGGQRLWGLGNPLEPKETHEFFVKSQYUKVLGNIDVGK-DIAQLL 357
 QY 370 CGGGIAAUGXPYLPQTVGVVJQKMTIAKEELFGVPMJILKFKTIEEVVGKANNSTYGLA 429
 DB 358 TCGVRKGDKGFAEPTIEVNEKPGSKIWFEBIFGPNLSIKTFKTEEAETANDTYGLA 417
 QY 430 AAVFTKDLKANYLSQALQAGTVMVNYVPSAUSPPRNVKMSGSHPELSEYGLDAYTPV 474
 DB 418 SVYTKSLNRLVRSVALETGGVSNFPPETPTPTPGMKQSGSGRELGEGLKAYLEP 477
 QY 490 KTVTKV 496
 DB 478 KTIINH 484

RESULT 12
 US-09-134-001C-4246
 ; Sequence 4246, Application US/09/34001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4246

; LENGTH: 488
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4246
 Query Match 28.7%, Score 753; DB 4; Length 488;
 Best Local Similarity 35.8%, Pred. No. 1.1e-71;
 Matches 173; Conservative 96; Mismatches 178; Indels 34; Gaps 11;
 QY 24 FINNEHDAVSRIKFTVNPSTGVICQVAGDKEDVDKAREGRGAFOLGSGWRRMDAS 83
 DB 21 YINGEWVDSASGETIDVINPATIEVMGKIAGNEEDVKNKAVD--AADKVYLEFRSSVE 77
 QY 84 HSGRLNPLADLIERDRTYLALETLDNGKPYVSYLVLDLDMVZLKLRYYAGWAPKYGK 143
 DB 78 REPELLIKIVKYEYQNPKNMCIFAITDEIGAPLSVSPNVHYCM---GNHPTAAPD----- 129
 QY 144 TIPTINGDFFSYTR-----HPPVGVCGGQITPWNFPILMQAWKLGCPALATGNVVMKV 197
 DB 130 --ALDSFOFEQRGDDLVVKEAIGVAGLVTPWNPFTNOTSIKLAFAAAGSPVVLKPSSE 187
 QY 198 TPLTALYVANLIKAEAGFPFGVGVNIVFSGPTAGAAASHEDVDKVAFTSTGTEGRVI 257
 DB 188 TFFAAIHILAEIFDKVGPKEGVNIVNGDGSVGNPLSEHPKVRMMSTGSGPTCSKIMEK 247
 QY 258 AGSSNLKRVTLGKGSNI'IMSADMDWAVEQAHPALFNU-QGCCACSGKRVKQ- 317
 DB 248 A-AKDFKVSLELGKSPYIVLDVDVFEAANATTKVNVNTGAVCTAGTRVILPSTIK 406
 QY 318 FEVFSVAPAKSPV-VGNPFSKTEHQCPVLEIGKFF IIGVINTGKPPAKIL 464
 DB 307 DYLI-AVREAFSKVKGQPEESTQVSPITSSKKQFLQVQYICKGINSAGLEYGSGKPK 365
 QY 373 -GIAADRGYFQTPVFGVDQGMIIAKEELFGVPMJILKFKTIEEVVGKANNSTYGLAA 431
 DB 366 EGL--DKGYFAPRTIFINVDNHMTIAOBEITFGPVMSVITYNNLDEATEIANDTKYGL 423
 QY 432 VETKEDKANYLSQALQAGTVMVNYVPSAUSPPRNVKMSGSHPELSEYGLDAYTPV 489
 DB 424 VICKIKITLPHVASTFACTTETIN--PAGKPTIDPQGVKFSQGLWVPSGLIVQW 480
 QY 490 KTV 492
 DB 481 KSI 483
 RESULT 13
 US-09-134-001C-4288
 ; Sequence 4288, Application US/09/14001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4388
 ; LENGTH: 493
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4388
 Query Match 25.9%, Score 681; DB 4; Length 493;
 Best Local Similarity 33.1%, Pred. No. 5.6e-64;
 Matches 159; Conservative 97; Mismatches 201; Indels 24; Gaps 7;
 QY 17 EVF--CNJFINNEHDAVSRIKFTVNPSTGVICQVAGDKEDVDKAREGRGAFOLG 74

```

Db      16  EVFNTNQLFINNEFISOSKETMDVNPATGSAFUIIILATEEVNDAIERSQA-QU- 73
QY      75  SPWREMGASHGRLNMIALIEPOTYLAALFTILN-KPYVIVSYLVGLIMVLRKLPYYA 134
Db      74  -EMERVQPTRAERHVKLLIPLEKRNDEIAQLYVKEQKTLAQAY-GEIDKSLISLIDYMI 131
QY      135 GWADKYHCK - ----TIPIDGDFPSYTRHEPVSQGLIPWNPFLIMQAWKLSPALA 185
Db      132 SLSMSDKGRVLQNSIANETIQI-----INPKIGVTAGIVPWNAPILVLMRKVIPALV 183
QY      186 TGNVVMKVAEOPPLTALYVANLIKPAEPPPGVNNIVPGFGPTAGAAATASHEDVDKVAFT 245
Db      184 TGSVWIKPSEETILLTLRLAELFRASITIPAGLFGIVPGTGETVGTQLASHKDIQLISIL 243
QY      246 GSTEIGRVIOVAGSSNKKVITLGLCKSPNTIMSDADMDWAVEGAHFAIFNQGCCCA 305
Db      244 GSRAGKSVYENA-AQTVKKNVLELGNAPVIVTSNADLLKAVNYIVTARINNAGQVCTC 302
QY      306 GSRTFVEDIVDFEVVRSARAKSRVVGPNPFSKTPQGPQVDPTQPKTILGYINTSKQFG 465
Db      303 PERLFVEDVIDDELNKVTSKMKSLVQGPFGENTGYGALINQKULDSHEKVQAIAKNG 462
QY      366 AKLIGGRIADRYFIQPTVFGVQDGMTIAKEELFGPVMQILKEFTIEEYVGRANNST 425
Db      363 ATLMTGGHQI KPHGFEYAPTVLDNRKDYNEKDEIFGDLVLAITTYRDFEQVIEDANDTN 422
QY      426 YGLAAAVFTKDLKANVLSQAQSTVWVNCYDVFCAQSPFGGKYMKSQSGRELGEYGLQA 485
Db      423 AGLSSYIFSENLTETVMTATERLKEFVYANCEREEVYNGYHAGWRRESGLGGADGTHGFE 482
QY      486 Y 486
Db      483 Y 483

```

RESULT 14

```

US-09-134-001C:4451
; Sequence 4451, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Dourcotte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 06/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C:4451

```

Query Match 24.8%; Score 652; DB 4; Length 518;

Best Local Similarity 34.9%; Pred No 7 6e-61;

Matches 169; Conservative 82; Mismatches 209; Indels 24; Gaps 15;

```

QY      23  IFINNEHJAVSRKTFETWNP-STGEVICQVAREGKEDVDKAREGRPGAFQSGSWRRMD 81
Db      43  LVINGE--KLTKTDDTNSNPANTSQILAKVSKATQDDIETAKAFESANHAYOSWRKWSHKD 100
QY      82  ASHSGRLNLKLIADIERDRITYLAALFTLNGKRPVIVSYLVLDLDMVLRKYA-GWADKY 140
Db      101  RA---ELLKVAALIRREKEEISAINVEAGKPV-DEAVGGAEGGIDETIYAKSMELA 156
QY      141  HGK-TIPIDGDFPSYTRHEPVSQGLIPWNPFL-LMQAWKLGALATGNTVVVWAKVAEQT 198
Db      157  DCKPVLDRGEHNRYP-YKPIGTGVIPWNPFPALMAGTTLAPVVA-GNTVLLKPARDT 214

```

```

QY      199  FLIALYVANLIKPAEPPPGVNNIVPGFGPTAGAAATASHEDVDKVAFTASTFEG 254
Db      215  VLTAYLMELEEAELFGPVMQILKEFTIEEYVGRANNST 274
QY      254  IYVAASSNKKRVITLGLCKSPNTIMSDADMDWAVEGAHFAIFNQGCCCA 414
Db      275  AVVFEQGLFKPVLAHMEKQIAIVNNVNDLILAAALIVISAGLSQKCSASNAIVDQ 434
QY      314  DIYDEFVVSVAAPKSRVVSNPFSKTPQGPQVDPTQPKTILGYINTSKQPKAKTIDGAS 474
Db      335  DVHDEILEKALQITOKLTLCNT-EENTFMGIVINOKOFDKIKNYTIFGKKEG KLEIGGG 492
QY      374  JAAIKRYPIDPTVFGVQDGMTIAKEELFGPVMQILKEFTIEEYVGRANNSTYGLAAAVF 444
Db      393  TLDSTGYFTPTFSLQSAIPIMQBEITEGVVGTIKVKDFDEATEVANDIYGLIGAVI 452
QY      434  TKDLKANVLSQAQSTVWVNCYDVFCAQSPFGGKYMKSQSGRELGEYGLQAYTEV 489
Db      453  TNHPEHWIKAVNFEEDVNIYNGSTIAAVVYHDEFGGKMSKSLIAKFGSDY LNFLEQ 511
QY      490  KTVT 494
Db      512  KVS 515

```

RESULT 15

```

US-09-155-183-4
; Sequence 4, Application US/09155184
; Patent No. 6324011
; GENERAL INFORMATION:
; APPLICANT: Natbad, Atjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,184
; CURRENT FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

Query Match

Best Local Similarity 22.9%; Score 692; DB 4; Length 482;

Matches 153; Conservative 94; Mismatches 199; Indels 24; Gaps 11;

```

QY      36  KTEFTVNPSTGEVICQVAREGKEDVDKAREGRPGAFQSGSWRRMDAS HSEGLINKLAD 94
Db      19  RTFERENPVTVGELVSPVAAATLEDAVAVAAGQAF---PAAWALAPNERSKELKAAEQ 75
QY      95  LIEPOTYLAALFTLNGKRPVIVSYLVLDLDMVLRKYA-GWADKYHCKTIPD GD 150
Db      76  LQARSGEITEA-----GETGAMANNYGNVILANMLKEAASMTYGVNGVITSDVIGRS 140
QY      151  FFSYTRHEPVSQGLIPWNPFL-LMQAWKLGALATGNTVVVWAKVAEQTILYAVANLIR 210
Db      131  PAAALR-OPCGVIGIAPWNAIVILATRALAMTACGNITVILKASELSIAVIRLIGVLO 189
QY      211  EAGFPVAVVNVINGEFGPTACAAIA-----SHEHIVKVAFTGSTEIDPTLVVAGSSNKKRV 266
Db      190  DAGIDGSGVNVVTSN APATAAGIVETIANTAVPVPNFTSTTHVGRVIG FLISAPH KPA 247
QY      267  TLFGCKSPNTIMSDADMDWAVEGAHFAIFNQGCCCAKSTFVQDILYDFEVVRSVAR 426

```


Db 248 LLEUGKAPILLVITQADLEAAVGAARFAYFNQICMSTEPILIVDAKVADAFVAGLAAR 307
QY 327 AKSRVVGNPDSKTEQGPQVDETQFKKILGY INIGKQEGAKILLGGGIAADRGYFIQPTV 386
Db 308 VETI FADUPADPESWIASU VITASAGTPIKALIDUAVAKGARI VIGGQL ---EGSTIQPTL 364
QY 367 FGDVGGMTIAKEETIRGPMGIIILKFTIEEVVGHANNISYGLAAAVETKDLKANYLSUA 446
Db 365 LGGVIAKRIYREESFGVAVVLNGEETEFALDGLANISRFPSAAIFSWTGPALALAP 424
QY 447 LQACTVWNVYGVFG-AUSPPGRYKMSGGRELGEYGLQAYTEVKIVTK 495
Db 425 VESGICHINGPTVHDEAUMPEGGVKSSGGSGGSGGKASIEHFQLRWVTIQ 474

Search completed: June 24, 2003, 10:31:40
Job time : 15.8469 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 ; Search time: 24.1632 seconds
(without alignments)
2938.528 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAATGAVPAPNCPVEVC... YGICAVIEVIVIKVYKRN 500

Scoring table:

BLOSUM62

Gapop 10 0 Gapext 0 5

Searched: 417779 seqs, 108206813 residues

417779

Total number of hits satisfying chosen parameters:

Minimum hit seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%

Database: Published_Applications_AA*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/BCI_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/MS06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
9: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
11: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
13: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1771	67.5	412	10	US-09-925-302-736
2	1778	65.4	512	9	US-09-961-403-12
3	1718	65.4	512	9	US-10-268-518-2
4	1718	65.4	512	9	US-10-205-823-14
5	1543	58.8	538	9	US-09-344-882-20
6	1543	58.8	538	9	US-10-293-882-20
7	1473	56.1	534	9	US-09-344-882-24
8	1473	56.1	534	9	US-10-293-882-24
9	1305.5	49.7	496	9	US-09-847-208-59
10	1280	48.8	501	9	US-09-344-882-22
11	1280	48.8	501	9	US-10-293-882-22
12	1254	46.3	495	9	US-09-847-208-11
13	1214	46.2	314	10	US-09-925-300-1348
14	1201.5	45.8	492	9	US-10-268-518-4
15	1191	45.4	493	10	US-10-175-906-21
16	1191	45.4	493	10	US-09-823-901-9
17	1128	43.0	500	9	US-10-166-087-4
18	978.5	37.3	490	16	US-09-815-242-1057
19	975.5	37.2	496	10	US-09-815-242-10550

20	947.5	36.1	496	10	US-09-815-242-12102	Sequence 1, App1
21	947.5	36.1	496	10	US-09-815-242-5644	Sequence 5644, Ap
22	947.5	36.1	496	10	US-09-815-242-12567	Sequence 1, App1
23	933.5	35.6	518	9	US-09-919-039-143	Sequence 143, Ap
24	917.5	35.0	508	1	US-08-781-966A-5241	Sequence 5241, Ap
25	914.5	34.8	506	9	US-09-738-626-6572	Sequence 6572, Ap
26	823	31.4	487	9	US-10-175-906-17	Sequence 17, App1
27	823	31.4	487	10	US-09-847-208-59	Sequence 5, App1
28	818.5	31.2	487	9	US-09-847-208-59	Sequence 3588, Ap
29	759.5	28.9	490	10	US-09-815-242-10574	Sequence 10574, A
30	759	28.9	481	10	US-09-815-242-10574	Sequence 12375, A
31	756	28.7	481	10	US-09-815-242-10574	Sequence 14047, A
32	754	28.7	475	10	US-09-815-242-10574	Sequence 5058, Ap
33	753	28.7	482	10	US-09-815-242-10574	Sequence 4108, Ap
34	695.5	26.5	497	9	US-09-738-626-4108	Sequence 3, App1
35	695.5	26.5	491	10	US-09-497-664-3	Sequence 6482, Ap
36	662.5	25.2	521	9	US-09-738-626-6482	Sequence 12, App1
37	654.5	24.9	498	9	US-10-272-419-12	Sequence 189, App
38	630	24.0	498	9	US-09-925-302-736	Sequence 6450, App
39	629	24.0	484	9	US-09-738-626-6450	Sequence 28, App1
40	621.5	23.7	481	1	US-08-976-0640-28	Sequence 1341b, A
41	606.5	23.1	481	1	US-09-738-626-4108	Sequence 5298, Ap
42	582.5	22.9	482	10	US-09-815-242-10574	Sequence 4037, Ap
43	548.5	20.9	479	10	US-09-815-242-1341b	
44	540	20.5	192	9	US-10-105-698-5298	
45	522.5	20.3	453	9	US-09-738-626-4037	

ALIGNMENTS

RESULT 1
US-09-925-302-736
Sequence 736, App1
Patent No. US-2002004941A1
GENERAL INFORMATION:
APPLICANT: Posen et al.
FILE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: FA04
CURRENT APPLICATION NUMBER: US-09-925-302
CURRENT FILING DATE: 2001-08-10
PCT APPLICATION NUMBER: PCT/US00/05918
PCT FILING DATE: 2000-03-08
LOCAL APPLICATION NUMBER: 62/214,275
PRIORITY FILING DATE: 1994-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 736
LENGTH: 412
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-302-736

Query Match 67.5% Score 1771; DR 10; Length 412;
Best Local Similarity 90.0% Pred. No. 6,296,156;
Matches 393 Conservative 7 Mismatches 293 Indels 0; Gaps 0;
1 MSAATGAVPAPNCPVEVCNCTINNEHFAVEKPTPVNSTEVICGVESKRY 59
|||||
20 USAAAGAVPAPNCPVEVCNCTINNEHFAVEKPTPVNSTEVICGVESKRY 79
|||||
61 FRAPDPAATFADDSWFMKASBSQELNLAKLALGKFLAATETGKPPVYEVY 120
|||||
80 FAVVAVAVALDELSTWFMKASBSQELNLAKLALGKFLAATETGKPPVYEVY 139
|||||
121 VGLGAVKLYAYAWAKRYKRTFTGQVSTSTREPPVAVGQITDMMPTIMAWCI 180
|||||
140 VGLGAVKLYAYAWAKRYKRTFTGQVSTSTREPPVAVGQITDMMPTIMAWCI 199
|||||
181 GALATGVVVVAKVAEPTFALVAVANIKIAGFPVAVNIINDEFTATAIASHIVV 240
|||||
200 GVALAGVVVAKVAEPTFALVAVANIKIAGFPVAVNIINDEFTATAIASHIVV 259
|||||


```

: APPLICANT: Monahan, John E.
: APPLICANT: Endege, Wilson O.
: APPLICANT: Ganavarepu, Manjula
: APPLICANT: Gorbacheva, Bella
: APPLICANT: Hoeisch, Sebastian
: APPLICANT: Kamatkar, Subhang
: APPLICANT: Womsey, Angela M
: APPLICANT: Glatt, Karen
: APPLICANT: Zhao, Xumei
: APPLICANT: Anderson, Justin
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
: TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
: FILE REFERENCE: MRI-044
: CURRENT APPLICATION NUMBER: US/10/205,823
: CURRENT FILING DATE: 2002-07-25
: PRIOR APPLICATION NUMBER: 60/307,982
: PRIOR FILING DATE: 2001-07-25
: PRIOR APPLICATION NUMBER: 60/414,456
: PRIOR FILING DATE: 2001-08-22
: PRIOR APPLICATION NUMBER: 60/335,020
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 60/341,746
: PRIOR FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: 60/362,158
: PRIOR FILING DATE: 2002-03-05
: NUMBER OF SEQ ID NOS: 455
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 14
: LENGTH: 512
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-205-823-14

```

```

Query Match 65.48, Score 1718, DR 2, Length 512
Best Local Similarity 65.58, Pred. No 7 qe-151;
Matches 323; Conservative 66; Mismatches 104; Indels 0; Gaps 0;

QY 8 AVPAPOQPEVFCNQIFINNEHUVASRKIFPTVNPSTGEVTCQVAREKDEVDKAREP 67
Db 19 ALPRIRNLVVKFKIFINNEHESKSKKFKATCNSTREQICEVEEGDKPDVKAREAA 78
QY 68 PGATGASPPMPASHSGPLNPLADLIERDTYLAALLETLDNGKPYVSYLVLDVYL 127
Db 79 QVAFQSPSPRPLNALSFGRLHLQDLADLVERIDATLAALLETMDTKPFLHAFIDLEGCI 138
QY 128 KCLRYAGWADKYHGKTIPTDGSFYSYTRHPVGVGGQIIPNFPILMCAWKLPALATG 187
Db 139 RLYFAGWADKIQGKIPTIDDDVWCFTRHPVGVGGQIIPNFPILMCAWKLPALATG 198
QY 188 NVVYKVAEOTPLTALYVNLKEAGFPPGVVNVIPGFGPTAGAAATASHEDYDKVFTGS 247
Db 199 NTMVLKPAEOTPLTALYVNLKEAGFPPGVVNVIPGFGPTAGAAATASHEDYDKVFTGS 258
QY 248 TEIIEVTCVAGSSNLKEVTELEGGKSFNINSLAUMENWAVEAHEAFENAGQ 307
Db 259 TEVGLVKEAASNLKRVTELEGGKSFNINSLAUMENWAVEAHEAFENAGQ 318
QY 308 RYFVQEDIDYEFVVRVARAKSVVNPFSKTEQGQVQVDEIQFKILGYINTGQEGAK 367
Db 319 RVFVEEQVYSEFVPSRSEYAKKRPVGDPEVDVTEQGPQIQKGFLELIESGKKEGAK 378
QY 368 LLOGRTAAAGYFQPTVPLVQJGMIIAKEEIPGVVNLKFKTEIEVVGGRANNIYG 427
Db 379 LECGGSAMEDKGLFKTPVSEVTDNNRIAKEEIPGVVNLKFKTEIEVVGGRANNIYG 438
QY 428 LAAAVFTKDLKANYLSUALQAGTVVWVYDVFCASPPGCKYKMSGSGRELGEY 487
Db 439 LTAAVFTKDLKANYLSUALQAGTVVWVYDVFCASPPGCKYKMSGSGRELGEY 498
QY 488 EVKTVTVKVPQKN 500
Db 499 EVKTVTVKVPQKN 511

```

```

RESULT 5
US-09-344-882-20
: Sequence 20, Application US/09344882
: Patent No. US 20020162137A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schnable, Patrick S
: APPLICANT: Ke, Jinshan
: APPLICANT: Johnson, Jerry L
: APPLICANT: Allred, Carolyn C
: APPLICANT: Fatland, Beth
: APPLICANT: Lutziger, Isabelle
: APPLICANT: Wee, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: TITLE OF INVENTION: Acetyl CoA Levels in Plants
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09/344,882
: CURRENT FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: US 60/090,717
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn Ver. 2.2
: SEQ ID NO 20
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-09-344-882-20

```

```

Query Match 58.88, Score 1543, DB 9, Length 538;
Best Local Similarity 61.14, Pred. No. 1 se-134;
Matches 300; Conservative 61; Mismatches 124; Indels 4; Gaps 2;

QY 2 SAAATQAVPAPNOQPEVFCNQIFINNEHUVASRKIFPTVNPSTGEVTCQVAREKDEVD 61
Db 42 SAAAREEII---NPSVQVSHQTQLLINNFVDSASGKTFTPTDPTGEVIAHVAEGDAEDIN 98
QY 62 KAREGHPHAPQJASPPMPASHSGPLNPLADLIERDTYLAALLETLDNGKPYVSYLV 121
Db 99 PAVKAAFTATPTFG-PMPKMSAYERSVLLRFADLVEKHESEELASLETWDNGKPYQOOLTA 157
QY 122 DLMVLKCLRYAGWADKYHGKTIPTDGSFYSYTRHPVGVGGQIIPNFPILMCAWK 181
Db 158 EIPWFAPLPFYASGAWKTHGLTTPAGVNYVHTLHEPFGVAGQIIPNFPILMFAHVG 217
QY 182 PALATGNVYVYKVAEOTPLTALYVNLKEAGFPPGVVNVIPGFGPTAGAAATASHEDYDK 241
Db 218 PALACGNTIVKTAFATPTAFYAGKLFLEAAGLPTVIVIVSGFATAAALASHMUVDK 277
QY 242 VATGSTEIGRPVTCVAGSSNLKEVTELEGGKSFNINSLAUMENWAVEAHEAFENAGQ 301
Db 278 LAFTHSTPTSKVILGRLAANSN:KPTVIEGKSPVIVFEADITKRAFTAHAFENAGQ 337
QY 302 CCAGSPTFPVQFQIYHFFVVPVSPVAPAKSVVNPFSKTEQGQVQVDEIQFKILGYINTG 361
Db 448 CCAGSPTFPVHVKVYSEFVPSRSEYAKKRPVGDPEVDVTEQGPQIQKGFLELIESGK 397
QY 362 KQEAALGAGYFQPTVPLVQJGMIIAKEEIPGVVNLKFKTEIEVVGGRANNIYG 421
Db 398 IESNATLECGDQIGKGYFQPTVFSNVKDDMLIAQDRIFGVGSILKFSDDVDEVIKRA 457
QY 422 NNSTYGLAAVFTKDLKANYLSUALQAGTVVWVYDVFCASPPGCKYKMSGSGRELGEY 481
Db 458 NETRYGLAAVFTKDLKANYLSUALQAGTVVWVYDVFCASPPGCKYKMSGSGRELGEY 517
QY 482 GLQAYTEVTKV 492
Db 518 SLNNYLIQIKAV 528

```

RESULT 6

US-10-293-865-20
 : Sequence 20, Application US/10293865
 : Publication No. US20030106090A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Wurtelo, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinshan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Allred, Carolyn C
 : APPLICANT: Falland, Beth
 : APPLICANT: Lutziqer, Isabelle
 : APPLICANT: Wen, Tsui-Jung
 : TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 : TITLE OF INVENTION: Acetyl CoA Levels in Plants
 : FILE REFERENCE: 217113
 : CURRENT APPLICATION NUMBER: US/10293, 865
 : CURRENT FILING DATE: 2002-11-13
 : PRIOR FILING DATE: 1999-06-25
 : PRIOR APPLICATION NUMBER: US 09/344,882
 : PRIOR FILING DATE: 1998-06-26
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: PatentIn Ver. 3.1
 : SEQ ID NO 20
 : LENGTH: 538
 : TYPE: PRT
 : ORGANISM: Arabidopsis Thaliana
 US-10-293-365-20

Query Match 58.8% Score 1543; DB 9; Length 538;
 Best Local Similarity 61.1% Pred. No. 1,5e+134;
 Matches 300; Conservative 61; Mismatches 126; Indels 4; Gaps 2;

QY	2	SAAATQAVPAPNQOFEVFCNQIFINNEWHDAVSRTFTVPNPSTGEVLCQVAEGDKEDVO	61
DB	42	SAAAEELI--NPVQVSHSTQLLINGNEVDSASGKTFTPLDPTGEVIAHVAEGDAEDIN	98
QY	62	KAREGRGAFQASGPPMPPDASHSGPLINPLADLIERUKTYLAALETLNDCKPYVSYLV	121
DB	99	RAVKAATAFDEG-PMPKMSAYERSVLLRFADLVEKHSSEELASLETWUNGKPYQSLTA	157
QY	122	DLDMVLKCLRYAGWADKYHGKTIPTDGDFFSYTRHEPVGCGQITPWNFPLLMQAWKLG	181
DB	158	EIPMFALPRYYAGWAIKTHGLTIPADGNYQVHTLHEPIGVAGQITPWNFPLLMQAWKLG	217
QY	182	PALATGNVVMKVAHQTPLTALYVANLKEAGFPFGVNVNIVPGFGPTAGAAIASHEDVDK	241
DB	218	PALACGNTIVLKTAEQTPLTAFYAGKLFLEAGLPPGVNLIVSGFGATAGALASHMDVDK	277
QY	242	VAFGTSGTIGRVTQVAGSSNLKRVTLLEGGKSPNIMSDALMDWAVEQAHFALFENQO	301
DB	278	LATGSTDGKVLILGAANSNLKPVLELGGKSPFIVFEDADIDKAVELAHFALFENQOQ	337
QY	302	CCAGSKRTFVOEDIDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	361
DB	338	CCAGSKRTFVHEKVDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	397
QY	362	KQEKAKLGGGRTAAADPGYFIQPTVPVGVQVQMTAKFEIFCPVMQLKFKTIFEVVGR	421
DB	398	IESNATLEGGDQIDGDKGFYFIQPTVPVGVQVQMTAKFEIFCPVMQLKFKTIFEVVGR	457
QY	422	NNSYGLAAAVFTIKDLKANYLSQALQAGTVWVNCYDFVGAQSPGPGYKMSGSGRELGEY	481
DB	458	NETKYGLAACVFTKNDTANRYSRALKAGTVWVNCYDFVGAQSPGPGYKMSGSGRELGEY	517
QY	482	GLOAYTEVKTV 492	
DB	518	SLNNYLOKAVVTSI 528	

RESULT 7

US 09 344 882 24
 : Sequence 24, Application US/09344882
 : Patent No. US20020162137A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Wurtelo, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinshan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Allred, Carolyn C
 : APPLICANT: Falland, Beth
 : APPLICANT: Lutziqer, Isabelle
 : APPLICANT: Wen, Tsui-Jung
 : TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 : TITLE OF INVENTION: Acetyl CoA Levels in Plants
 : FILE REFERENCE: 201573
 : CURRENT APPLICATION NUMBER: US/09/344,882
 : CURRENT FILING DATE: 1999-06-25
 : PRIOR APPLICATION NUMBER: US 60/090,717
 : PRIOR FILING DATE: 1998-06-26
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: PatentIn Ver. 2.2
 : SEQ ID NO 24
 : LENGTH: 534
 : TYPE: PRT
 : ORGANISM: Arabidopsis Thaliana
 US-09-344-882-24

Query Match 56.1% Score 1473; DB 9; Length 534;
 Best Local Similarity 59.3% Pred. No. 4.5e+128;
 Matches 294; Conservative 55; Mismatches 143; Indels 4; Gaps 2;

QY	1	MSAATQAVPAPNQOFEVFCNQIFINNEWHDAVSRTFTVPNPSTGEVLCQVAEGDKEDVO	60
DB	37	LAATAVENTITPP---VKVEHIGLLGGREVDVAVSKTEPTLDRNCEVIAVSEGDADVD	94
QY	61	DKAREGRGAFQASGPPMPPDASHSGPLINPLADLIERUKTYLAALETLNDCKPYVSYLV	120
DB	94	NAAVAARAKAFDGG-TWPKMTAYERSKTLERFADLIERKNDLIEKNDLIEKNDLIEKNDLIE	152
QY	121	VDFDQVVKI KTVAVAWAWVYKCTIPDGHDFSYTRHEPVGCGQITPWNFPLLMQAWKLG	180
DB	153	TEVPMILAKVFKYAW	212
QY	181	GPALATGNVVMKVAHQTPLTALYVANLKEAGFPFGVNVNIVPGFGPTAGAAIASHEDVDK	240
DB	213	GPALAGNTVILKTAEQTPLTAFYAGKLFLEAGLPPGVNLIVSGFGATAGALASHMDVDK	272
QY	241	KVAFGTSGTIGRVTQVAGSSNLKRVTLLEGGKSPNIMSDALMDWAVEQAHFALFENQO	300
DB	274	KVAFGTSGTIGRVTQVAGSSNLKRVTLLEGGKSPNIMSDALMDWAVEQAHFALFENQO	332
QY	301	CCAGSKRTFVOEDIDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	360
DB	334	CCAGSKRTFVHEKVDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	392
QY	361	KQEKAKLGGGRTAAADPGYFIQPTVPVGVQVQMTAKFEIFCPVMQLKFKTIFEVVGR	420
DB	394	GVEGATLQAGGDRGSKGYITQPTVPVGVQVQMTAKFEIFCPVMQLKFKTIFEVVGR	452
QY	421	NNSYGLAAAVFTIKDLKANYLSQALQAGTVWVNCYDFVGAQSPGPGYKMSGSGRELGEY	480
DB	453	NNSYGLAAAVFTIKDLKANYLSQALQAGTVWVNCYDFVGAQSPGPGYKMSGSGRELGEY	512
QY	481	GLOAYTEVKTV 496	
DB	513	YSNNYLOKAVVTSI 528	


```

: APPLICANT: Fatland, Beth
: APPLICANT: Lutziqer, Isabelle
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09/344,882
: PRIOR FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: Patent In Ver. 2.2
: SEQ ID NO 22
: LENGTH: 501
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

```

```

Query Match      48.8%  Score 1280; DB 9; Length 501;
Best Local Similarity 53.1%  Pred. No. 3.4e-110;
Matches 255; Conservative 74; Mismatches 149; Indels 2; Gaps 2;

QY 16 PEVPCNOIFINNEHDAVSRKTEPTVNPSTGEVICQVAGKEDVDKAREGRPGAFQLGS 75
Db 15 PEIKFTKLFGINGOFIDAAGSKTETIDPRNGEVIAIARGKEDVDLAVNAARYAFDHG- 73
QY 76 PWRRMDASHSGRLNRIADLIERDRTYLAALETLDNGKPYVISYLVLDLMDVLKCLRYVAG 135
Db 74 PWPRMTGFERAKLNLKFALEENIEELAKLDADVGKILFQLGKYADIPATAGHFRVYAG 133
QY 136 WAKYKRGKTIID-GDFFSYTHREIPVCGQIIPNFPFLDMQAMKLGHALATGNVVMKV 194
Db 134 AAKIHGETLKMTRQSLFGTYLKEPIGVVGNIPNPFISINEATKVAPEAMAAGTMMVKP 193
QY 195 ABQTPITATYVANLKEAGFPVGVNIVPGFPIACAAASHEDVDKVAFTGSTEIGRVI 254
Db 194 ABQISLSALFYAHLKSEAGIPGVGLNIVTGFGSTAGAAASHMDVDKVSFGSTGVGRKI 253
QY 255 QVAAAGSNLKRVTLELGGKSPNINSADMLWAVEQAIFALFENAGQCCAGSKTFVQED 314
Db 254 MQAAAASNLKVKVSLGLGKSPILLFNDAIDUKAADLALLGCFYNNKGEICVASSKVFVQEG 313
QY 315 IYDEFVVRVARAKSVGNPNDSKTEQGPQVDETFQPKKILGYINTGQEGAKILLCGGGI 374
Db 314 IYDKVVEKIVEKAKDWTGDDPDSARQGPQVDETFQPKKILGYINTGQEGAKILLCGGKA 373
QY 375 AADRGYFIQPTVFQVQDGMIIAKEEIPGVPMOILKFKTIEEVGSRANNSTYGLAAAVFT 434
Db 374 IGDYGYFIQPTIFALVTLMKLYJGEIDSPVMSLMKPEFVVEGLRANNIYKGLAAHLS 433
QY 435 KULDKANYLSQALQAGTVWVNYDVFAGSPRGGYKMSGRELGEYGLQAYTEKTVTV 494
Db 434 QDIDLINTVSRISIKAGITWVNYFGFDLDCPYGYKMSGNCRESGMALDNYLQTKSVVM 493

```

```

RESULT 11
US-09-344-882-22
: Sequence 22, Application US/10293865
: Publication No. US20030106090A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schnable, Patrick S
: APPLICANT: Ke, Jinsan
: APPLICANT: Johnson, Jerry L
: APPLICANT: Allied, Carolyn C
: APPLICANT: Fatland, Beth
: APPLICANT: Lutziqer, Isabelle
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: FILE REFERENCE: Acetyl CoA Levels in Plants

```

```

: FILE REFERENCE: 217114
: CURRENT APPLICATION NUMBER: US/10/293,865
: PRIOR FILING DATE: 2002-11-14
: PRIOR FILING DATE: 1999-06-25
: PRIOR FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: Patent In Ver. 3.1
: SEQ ID NO 22
: LENGTH: 501
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana
US-10-293-865-22

```

```

Query Match      48.8%  Score 1280; DB 9; Length 501;
Best Local Similarity 53.1%  Pred. No. 3.4e-110;
Matches 255; Conservative 74; Mismatches 149; Indels 2; Gaps 2;

QY 16 PEVPCNOIFINNEHDAVSRKTEPTVNPSTGEVICQVAGKEDVDKAREGRPGAFQLGS 75
Db 15 PEIKFTKLFGINGOFIDAAGSKTETIDPRNGEVIAIARGKEDVDLAVNAARYAFDHG- 74
QY 76 PWRRMDASHSGRLNRIADLIERDRTYLAALETLDNGKPYVISYLVLDLMDVLKCLRYVAG 135
Db 74 PWPRMTGFERAKLNLKFALEENIEELAKLDADVGKILFQLGKYADIPATAGHFRVYAG 134
QY 136 WAKYKRGKTIID-GDFFSYTHREIPVCGQIIPNFPFLDMQAMKLGHALATGNVVMKV 194
Db 134 AAKIHGETLKMTRQSLFGTYLKEPIGVVGNIPNPFISINEATKVAPEAMAAGTMMVKP 193
QY 195 ABQTPITATYVANLKEAGFPVGVNIVPGFPIACAAASHEDVDKVAFTGSTEIGRVI 254
Db 194 ABQISLSALFYAHLKSEAGIPGVGLNIVTGFGSTAGAAASHMDVDKVSFGSTGVGRKI 253
QY 255 QVAAAGSNLKRVTLELGGKSPNINSADMLWAVEQAIFALFENAGQCCAGSKTFVQED 314
Db 254 MQAAAASNLKVKVSLGLGKSPILLFNDAIDUKAADLALLGCFYNNKGEICVASSKVFVQEG 313
QY 315 IYDEFVVRVARAKSVGNPNDSKTEQGPQVDETFQPKKILGYINTGQEGAKILLCGGGI 374
Db 314 IYDKVVEKIVEKAKDWTGDDPDSARQGPQVDETFQPKKILGYINTGQEGAKILLCGGKA 373
QY 375 AADRGYFIQPTVFQVQDGMIIAKEEIPGVPMOILKFKTIEEVGSRANNSTYGLAAAVFT 434
Db 374 IGDYGYFIQPTIFALVTLMKLYJGEIDSPVMSLMKPEFVVEGLRANNIYKGLAAHLS 433
QY 435 KULDKANYLSQALQAGTVWVNYDVFAGSPRGGYKMSGRELGEYGLQAYTEKTVTV 494
Db 434 QDIDLINTVSRISIKAGITWVNYFGFDLDCPYGYKMSGNCRESGMALDNYLQTKSVVM 493

```

```

RESULT 12
US-09-847-208-11
: Sequence 11, Application US/09847208
: Publication No. US20030082190A1
: GENERAL INFORMATION:
: APPLICANT: Saxena, Androw
: APPLICANT: Zhang, Ke
: APPLICANT: Zhu, Daocunq
: TITLE OF INVENTION: Fusion Molecules and Treatment of
: FILE REFERENCE: IgE-MEDIATED ALLERGIC DISEASES
: CURRENT APPLICATION NUMBER: US/09/847,208
: CURRENT FILING DATE: 2001-05-01
: NUMBER OF SEQ ID NOS: 177
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11
: LENGTH: 495
: TYPE: PRT
: ORGANISM: Alternaria alternata
US-09-847-208-11

```


[illegible]

RESULT 15

```

US-10-175-696-21
; Sequence 21, Application US/10175696
; Publication No. US20030092658A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Meyers, Rachel
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: NOVEL HUMAN ENZYME FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-193001
; CURRENT APPLICATION NUMBER: US/10/175,696
; PRIOR FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 10/067,668
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/843,901
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: PCT/US01/10720
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/193,920
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/862,658
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16380
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/205,675
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/882,837
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19319
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/211,727
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
; S-10-175-696-21

```

[illegible]

Search completed: June 24, 2003, 10:37:05
Job time : 26.1692 secs

GenCore version 5.1.6
Copyright (c) 1994 - 2003, Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:11:00 ; Search time 141.017 seconds
(without alignments)
1780.867 Million cell updates/sec

Title: US-09-830-751-4
Perfect score: 2625
Sequence: 1 MSAAATCAVAPAPNQPEVFC YGLQAYTEKIVIKVPQKN 500

Scoring table: RUSUM62
Gap: 10 0, Gapext 0.5

Searched: 4569144 seqs, 64473110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Pending_Patent's_AA_Main.*
- 1: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
 - 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
 - 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
 - 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
 - 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
 - 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
 - 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
 - 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
 - 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
 - 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
 - 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
 - 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
 - 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
 - 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
 - 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
 - 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
 - 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
 - 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
 - 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
 - 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
 - 26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
 - 27: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2624	99.9	516	21	US-09-791-537-77606
2	2574	98.1	516	7	US-08-466-2108-4
3	2574	98.1	515	21	US-09-791-537-77611
4	2572	98.0	517	19	US-09-538-092-851
5	2572	98.0	517	21	US-09-791-537-74079
6	2572	98.0	517	27	US-09-389-987-1242

7	2572	98.0	517	27	US-09-412-418-1242
8	2570	47.4	516	7	US-08-466-2108-4
9	2569	97.9	517	21	US-09-791-537-71773
10	2569	97.9	517	27	US-09-389-987-56
11	2569	97.9	517	27	US-09-412-418-56
12	2544	95.4	494	21	US-09-791-537-99557
13	2503	95.4	500	21	US-09-791-537-64443
14	2500	95.2	516	21	US-09-791-537-74080
15	2480	94.5	519	21	US-09-791-537-130904
16	2447.5	93.2	519	21	US-09-763-443-1247
17	2447.5	93.2	519	21	US-09-760-476-3423
18	2447.5	93.2	494	21	US-10-212-054-1247
19	2436	92.8	494	21	US-09-791-537-19515
20	2436	92.8	520	21	US-09-791-537-74076
21	2434	92.7	500	10	US-08-512-853-13
22	2434	92.7	500	21	US-09-791-537-74077
23	2434	92.7	500	21	US-08-512-853-12
24	2413	91.9	517	21	US-09-791-537-76513
25	2083	74.4	494	21	US-09-791-537-71778
26	1943	74.0	517	27	US-09-389-987-1227
27	1943	74.0	517	27	US-09-412-418-1227
28	1939	73.9	517	21	US-09-791-537-50322
29	1933	73.6	559	21	US-09-760-443-1406
30	1933	73.6	559	21	US-09-760-476-3429
31	1933	73.6	559	26	US-10-212-054-1406
32	1929	73.5	517	1	PGT-US01-17253-3
33	1929	73.5	517	19	US-09-538-092-1047
34	1929	73.5	517	21	US-09-791-537-77611
35	1866	71.1	520	20	US-09-514-150-7212
36	1866	71.1	520	21	US-09-791-537-129281
37	1866	71.1	520	27	US-09-191-681-7243
38	1866	71.1	520	27	US-09-191-681-5649
39	1857	70.7	538	27	US-09-167-217-7335
40	1857	70.7	538	27	US-09-173-464-5881
41	1802	68.6	509	21	US-09-791-537-73639
42	1798	68.5	502	21	US-09-791-537-66277
43	1795	68.4	499	21	US-09-791-537-115118
44	1794	68.3	499	6	US-08-512-853-2
45	1794	68.3	499	21	US-09-791-537-12493

ALIGNMENTS

RESULT 1
US-09-791-537-77606
Sequence 77606, Application US/09791537

GENERAL INFORMATION:
APPLICANT: Biocompare, Inc.
APPLICANT: Debe, Derek
INVENTOR: Danzel, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001 02 22
NUMBER OF SEQ ID NOS: 154055
SOFTWARE: PatentIn version 3.0
SEQ ID NO 77606
LENGTH: 516
TYPE: PRT
ORGANISM: Homo sapiens
US-09-791-537-77606

Query Match 99.9%, Score 2622, DB 21, Length 516;
Best Local Similarity 99.8%, Pred. No. 5e-262;
Matches 499, Conserved 1, Mismatches 0, Indels 0, Gaps 0;

Q7	1	MSAAATCAVAPAPNQPEVFC	INQI	INNEWHDAVSHKIFIVNF	SEV	55	VAALGKEDV	60
DB	16	MSAAATCAVAPAPNQPEVFC	INQI	INNEWHDAVSHKIFIVNF	SEV	55	VAALGKEDV	75
Q7	41	DKAREFETGATLSTFWFPCASIS	TPLL	NPLACLTEPEFTYLAALFLD	WYV	120		

```
Db 76 DKAREGPGAFQCSWRMDASHGRLLNPLADLLERDTYLAALETLDCKKPVVSYL 135
QY 121 VLDLMLVKLYYAGWADKYHGKTIPIIDGDFTSYTRHEPVGVCGQIIPWNEPFLMQAWKL 180
Db 136 VLDLMLVKLYYAGWADKYHGKTIPIIDGDFTSYTRHEPVGVCGQIIPWNEPFLMQAWKL 195
QY 181 GPALATGNVVMKVAQTPLTALYVANLKEAGFPVGVVNIIPVGGPTAGAAIASHEDVD 240
Db 196 GPALATGNVVMKVAQTPLTALYVANLKEAGFPVGVVNIIPVGGPTAGAAIASHEDVD 255
QY 241 KVAFTGSTEIGRVQVAAAGSSNLKRVTLLELGCKSPNIMSUDMDWAVEQAHFALFFNQ 300
Db 256 KVAFTGSTEIGRVQVAAAGSSNLKRVTLLELGCKSPNIMSUDMDWAVEQAHFALFFNQ 315
QY 301 OCCAGSKTFVQEDYDFEYVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINT 360
Db 316 OCCAGSKTFVQEDYDFEYVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINT 375
QY 361 GKQGAKLLGGGTAADRGYFIQPTVFGDVQDGMTIAKEIIPGVNQILKEKTIPEVVGR 420
Db 376 GKQGAKLLGGGTAADRGYFIQPTVFGDVQDGMTIAKEIIPGVNQILKEKTIPEVVGR 435
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFQAGSPGCKYKMSGSGRELGE 480
Db 436 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFQAGSPGCKYKMSGSGRELGE 495
QY 481 YGLQAYTEVKTIVTKVPQKN 500
Db 496 YGLQAYTEVKTIVTKVPQKN 515
```

RESULT 2

```
US-08-466-210B-2
; Sequence 2, Application US/08366210B
; GENERAL INFORMATION:
; APPLICANT: Lauerman et al.
; TITLE OF INVENTION: Expression of a Mutant Form of Aldehyde
; TITLE OF INVENTION: Dehydrogenase as an Alcohol Aversive Agent
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,210H
; FILING DATE: December 29, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: TTI-128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-210B 2
```

Query Match 98.1%; Score 2574, DB 7, Length 515,
Best Local Similarity 99.8%; Pred. No. 4.8e-257;

```
Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 APNQOPEVFCNQIFINNEHDAVSKRTPTVNPSTGEVTCQVAEGIKELWIKAREGRGA 70
Db 25 APNQOPEVFCNQIFINNEHDAVSKRTPTVNPSTGEVTCQVAEGIKELWIKAREGRGA 84
QY 71 FOLGSPWRRMDASHGRLLNPLADLLERDTYLAALETLDCKKPVVSYLVLDLMLVKL 140
Db 85 FOLGSPWRRMDASHGRLLNPLADLLERDTYLAALETLDCKKPVVSYLVLDLMLVKL 144
QY 131 RYVAGWADKYHGKTIPIIDGDFTSYTRHEPVGVCGQIIPWNEPFLMQAWKLGFALATGNV 190
Db 145 RYVAGWADKYHGKTIPIIDGDFTSYTRHEPVGVCGQIIPWNEPFLMQAWKLGFALATGNV 204
QY 191 VMKVAQTPLTALYVANLKEAGFPVGVVNIIPVGGPTAGAAIASHEDVDKVAFTGSTEI 250
Db 205 VMKVAQTPLTALYVANLKEAGFPVGVVNIIPVGGPTAGAAIASHEDVDKVAFTGSTEI 264
QY 251 GRVIOVAAAGSSNLKRVTLLELGCKSPNIMSUDMDWAVEQAHFALFFNQGGTCAKSKTF 310
Db 265 GRVIOVAAAGSSNLKRVTLLELGCKSPNIMSUDMDWAVEQAHFALFFNQGGTCAKSKTF 324
QY 311 VQPIVYVFPVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINTQKQZAKL 370
Db 325 VQPIVYVFPVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINTQKQZAKL 384
QY 371 GGGTAADRGYFIQPTVFGDVQDGMTIAKEIIPGVNQILKEKTIPEVVGRKANNSTYGLAA 430
Db 385 GGGTAADRGYFIQPTVFGDVQDGMTIAKEIIPGVNQILKEKTIPEVVGRKANNSTYGLAA 444
QY 431 AVFTKLDKANYLSQALQAGTVMVNCYDVFQAGSPGCKYKMSGSGRELGEYGLQAYTEVK 490
Db 445 AVFTKLDKANYLSQALQAGTVMVNCYDVFQAGSPGCKYKMSGSGRELGEYGLQAYTEVK 504
QY 491 TVTVKVPQKN 500
Db 505 TVTVKVPQKN 514
```

RESULT 4

```
US-09-791-537-57517
; Sequence 4, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomimix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001 02 22
; NUMBER OF SEQ ID NOS: 154055
; SOFTWARE: Patent In version 4.0
; SEQ ID NO 57517
; LENGTH: 515
; TYPE: PRO
; ORGANISM: Homo sapiens
US-09-791-537-57517
```

Query Match 98.1%; Score 2574, DB 21, Length 515,
Best Local Similarity 99.8%; Pred. No. 4.8e-257;

```
Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 APNQOPEVFCNQIFINNEHDAVSKRTPTVNPSTGEVTCQVAEGIKELWIKAREGRGA 70
Db 25 APNQOPEVFCNQIFINNEHDAVSKRTPTVNPSTGEVTCQVAEGIKELWIKAREGRGA 84
QY 71 FOLGSPWRRMDASHGRLLNPLADLLERDTYLAALETLDCKKPVVSYLVLDLMLVKL 140
Db 85 FOLGSPWRRMDASHGRLLNPLADLLERDTYLAALETLDCKKPVVSYLVLDLMLVKL 144
QY 131 RYVAGWADKYHGKTIPIIDGDFTSYTRHEPVGVCGQIIPWNEPFLMQAWKLGFALATGNV 190
```

Db 145 PYYAGWADKYHSTKIPIINQDFSYVIRHEPVGVQSGIIPNNPPLIMLAWKLGALALGNVV 204
QY 191 VMKVAQPTLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKVAFQSSIRP 250
Db 205 VMKVAQPTLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKVAFQSSIRP 264
QY 251 GRVTVAAAGSSNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQSGVAVSRFP 410
Db 265 GRVTVAAAGSSNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQSGVAVSRFP 424
QY 311 VQEDIDYEFVVPVSAPAKSVWGNPFTSKTEQGEQVDETQFKILGYINTGQSGAKLLA 370
Db 325 VQEDIDYEFVVPVSAPAKSVWGNPFTSKTEQGEQVDETQFKILGYINTGQSGAKLLA 384
QY 371 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 430
Db 385 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 444
QY 431 AVFTKDLKDKANYLSOALQAGTVWVNCYDVFAGSPGKYKMSGSPPELAGYGIQAYTEVK 490
Db 445 AVFTKDLKDKANYLSOALQAGTVWVNCYDVFAGSPGKYKMSGSPPELAGYGIQAYTEVK 504
QY 491 TTVTKVPQKN 500
Db 505 TTVTKVPQKN 514

RESULT 4

US-09-538-092-851
; Sequence 851, Application US/09538092
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 138?
; SOFTWARE: GigaFats2/Formatter Version 0.9
; SEQ ID NO 851
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P05091
US-09-538-092-851

Query Match 98.0%, Score 2572, DB 19, Length 517,
Best Local Similarity 98.2%, Pred. No. 7.8e-257;
Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 1 MSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNVSTGEVICQVABGJKEIV 60
Db 17 LSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNVSTGEVICQVABGJKEIV 76
QY 61 DKAREGPGAFQSGSPWPPMDASHSGPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 120
Db 77 DKAVKAAPAAAFQSGSPWPPMDASHSGPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 136
QY 121 VLLDMVLKCLRYTAGWADKYHGKTIPIIDGDFSYTHREPVGVQSGIIPNNPPLIMLAWKLG 180
Db 137 VLLDMVLKCLRYTAGWADKYHGKTIPIIDGDFSYTHREPVGVQSGIIPNNPPLIMLAWKLG 196
QY 181 GPALATGNVVMKVAEUTPLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKV 240
Db 197 GPALATGNVVMKVAEUTPLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKV 256
QY 241 KVAFTQSTETGPRVLEVAASNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQ 300
Db 257 KVAFTQSTETGPRVLEVAASNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQ 316
QY 361 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 420
Db 437 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 444
QY 491 TTVTKVPQKN 500
Db 505 TTVTKVPQKN 514

RESULT 5

US-09-791-537-74079
; Sequence 74079, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biocomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74079
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-74079

Query Match 98.0%, Score 2572, DB 21, Length 517,
Best Local Similarity 98.2%, Pred. No. 7.8e-257;
Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNVSTGEVICQVABGJKEIV 60
Db 17 LSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNVSTGEVICQVABGJKEIV 76
QY 61 DKAREGPGAFQSGSPWPPMDASHSGPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 120
Db 77 DKAVKAAPAAAFQSGSPWPPMDASHSGPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 136
QY 121 VLLDMVLKCLRYTAGWADKYHGKTIPIIDGDFSYTHREPVGVQSGIIPNNPPLIMLAWKLG 180
Db 137 VLLDMVLKCLRYTAGWADKYHGKTIPIIDGDFSYTHREPVGVQSGIIPNNPPLIMLAWKLG 196
QY 181 GPALATGNVVMKVAEUTPLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKV 240
Db 197 GPALATGNVVMKVAEUTPLTALYVANLKEAGFPFGVNVIRPSPITAGAAIASHEWIKV 256
QY 241 KVAFTQSTETGPRVLEVAASNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQ 300
Db 257 KVAFTQSTETGPRVLEVAASNLKRPVTLLEGRKSPNIMSDALIMWAVELAHFALEFNGQ 316
QY 361 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 420
Db 437 GGGIAADPGYFIQPTVFGDVGQMTTAKBEIFSPVWQILKFKTIEFVWSPANNISYGLAA 444
QY 491 TTVTKVPQKN 500
Db 505 TTVTKVPQKN 514

QY 421 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 496
 QY 481 YGLQAYTEVKTVTVKVPQKN 500
 Db 497 YGLQAYTEVKTVTVKVPQKN 516

RESULT 6

US-60-389-987-1242
 ; Sequence 1242, Application US/6038987
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; CURRENT APPLICATION NUMBER: US/60/389,987
 ; CURRENT FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1242
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-389-987-1242

Query Match 98.0% Score 2572 DB 27 Length 517
 Best Local Similarity 98.2% Pred. No. 7.8e-257
 Matches 491 Conservative 2 Mismatches 7 Indels 0 Gaps 0

QY 1 MSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICQVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICQVAEGDKEDV 76
 QY 61 DKAREKPGAFQUGSPWRPMIAASHKRLNKLADLIERDTYLAALLETLDNGKPYVSYL 120
 Db 77 DKAVKAARAAAFQUGSPWRPMIAASHKRLNKLADLIERDTYLAALLETLDNGKPYVSYL 136
 QY 121 VOLDMVLKCLRYAYAGWADKYHCKTIPIDQDFESYTHHEPFWCGGIIHWNFPLLMQAWKL 180
 Db 137 VOLDMVLKCLRYAYAGWADKYHCKTIPIDQDFESYTHHEPFWCGGIIHWNFPLLMQAWKL 196
 QY 181 GPALATGNVVMKVAEQTTPTAIYVANLKEAGFPFGVNVIVPQPTAGAAIASHEIDV 240
 Db 197 GPALATGNVVMKVAEQTTPTAIYVANLKEAGFPFGVNVIVPQPTAGAAIASHEIDV 256
 QY 241 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFELGKSPNIMSDALMIAWAVEQAHEALFTNQ 400
 Db 257 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFELGKSPNIMSDALMIAWAVEQAHEALFTNQ 416
 QY 301 QCCAGSRFTVQEDIDYDEPVERSVARAKSRVGNFIPSKTEGPGVIGPQVITPKKIIIGYINT 360
 Db 317 QCCAGSRFTVQEDIDYDEPVERSVARAKSRVGNFIPSKTEGPGVIGPQVITPKKIIIGYINT 376
 QY 361 GKQEGAKLLCGGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVPMQILKEKTEIEVVGR 420
 Db 376 GKQEGAKLLCGGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVPMQILKEKTEIEVVGR 436
 QY 421 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 496
 QY 481 YGLQAYTEVKTVTVKVPQKN 500
 Db 497 YGLQAYTEVKTVTVKVPQKN 516

RESULT 7

US-60-412-418-1242
 ; Sequence 1242, Application US/60412418
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; CURRENT APPLICATION NUMBER: US/60/412,418
 ; CURRENT FILING DATE: 2002-09-20
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1242
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-412-418-1242

Query Match 98.0% Score 2572 DB 27 Length 517
 Best Local Similarity 98.2% Pred. No. 7.8e-257
 Matches 491 Conservative 2 Mismatches 7 Indels 0 Gaps 0

QY 1 MSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICQVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICQVAEGDKEDV 76
 QY 61 DKAREKPGAFQUGSPWRPMIAASHKRLNKLADLIERDTYLAALLETLDNGKPYVSYL 120
 Db 77 DKAVKAARAAAFQUGSPWRPMIAASHKRLNKLADLIERDTYLAALLETLDNGKPYVSYL 136
 QY 121 VOLDMVLKCLRYAYAGWADKYHCKTIPIDQDFESYTHHEPFWCGGIIHWNFPLLMQAWKL 180
 Db 137 VOLDMVLKCLRYAYAGWADKYHCKTIPIDQDFESYTHHEPFWCGGIIHWNFPLLMQAWKL 196
 QY 181 GPALATGNVVMKVAEQTTPTAIYVANLKEAGFPFGVNVIVPQPTAGAAIASHEIDV 240
 Db 197 GPALATGNVVMKVAEQTTPTAIYVANLKEAGFPFGVNVIVPQPTAGAAIASHEIDV 256
 QY 241 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFELGKSPNIMSDALMIAWAVEQAHEALFTNQ 400
 Db 257 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFELGKSPNIMSDALMIAWAVEQAHEALFTNQ 416
 QY 301 QCCAGSRFTVQEDIDYDEPVERSVARAKSRVGNFIPSKTEGPGVIGPQVITPKKIIIGYINT 360
 Db 317 QCCAGSRFTVQEDIDYDEPVERSVARAKSRVGNFIPSKTEGPGVIGPQVITPKKIIIGYINT 376
 QY 361 GKQEGAKLLCGGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVPMQILKEKTEIEVVGR 420
 Db 376 GKQEGAKLLCGGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVPMQILKEKTEIEVVGR 436
 QY 421 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGYKMSGSGRELGE 496
 QY 481 YGLQAYTEVKTVTVKVPQKN 500
 Db 497 YGLQAYTEVKTVTVKVPQKN 516

RESULT 8

US-08-466-2108-4
 ; Sequence 4, Application US/084662108
 ; GENERAL INFORMATION:
 ; APPLICANT: Lauterman et al.
 ; TITLE OF INVENTION: Expression of a Mutant Form of Alcoholdehyde
 ; TITLE OF INVENTION: Dehydrogenase as an Alcohol Acetate Agent

: NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LAHIVE & CUCKFIELD
 : STREET: 60 State Street, suite 510
 : CITY: Boston
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02109 1875
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: ASCII Text
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/366,210B
 : FILING DATE: December 29, 1994
 : CLASSIFICATION: 514
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Jane E. Remillard
 : REGISTRATION NUMBER: 38,872
 : REFERENCE/DOCKET NUMBER: TT1-128
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617)227-7400
 : TELEFAX: (617)227-5941
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 515 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-366-210R-4

Query Match 97.9%, Score 2570, DB 7, Length 515,
 Best Local Similarity 99.6%, pred. No. 1.3e-256;
 Matches 488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 11 APNOQPEVFCNOIFINNEWHDAVSPKTFPTVNPSTGEVTCQVAEGRKENVKAPGPPGA 70
 DB 25 APNOQPEVFCNOIFINNEWHDAVSPKTFPTVNPSTGEVTCQVAEGRKENVKAPGPPGA 84
 QY 71 FQAGSPWPMIAASHGPIINPLADLIEPDTYLAALFTLNGKPYVVISLVLDWVKL 130
 DB 85 FQAGSPWPMIAASHGPIINPLADLIEPDTYLAALFTLNGKPYVVISLVLDWVKL 144
 QY 131 RYAGWAHKKYHKTPTPTGTFEFSYTPRHPVSVVGGTTPWNEPLMLKAWKLGPAALGNNV 190
 DB 145 RYAGWAHKKYHKTPTPTGTFEFSYTPRHPVSVVGGTTPWNEPLMLKAWKLGPAALGNNV 204
 QY 191 VMKVAFTPTPTALYVANIIEKAGPPGVVNIIVPGFPTAGAAATASHETVGVKVAFTSTEL 250
 DB 205 VMKVAEQPTPLALYVANIIEKAGPPGVVNIIVPGFPTAGAAATASHETVGVKVAFTSTEL 264
 QY 261 GPVTCVAAGSSNIPKVTLEFGSKSFNINSLALMIMWAVEKAHFALEFNGQCCAGSPLE 310
 DB 265 GPVTCVAAGSSNIPKVTLEFGSKSFNINSLALMIMWAVEKAHFALEFNGQCCAGSPLE 324
 QY 311 VQEDYIGSEFVSVVAVAKSRVWVNPUSKTEQSPQVIEHLPKKIIGYINTKGRAKIL 370
 DB 325 VQEDYIGSEFVSVVAVAKSRVWVNPUSKTEQSPQVIEHLPKKIIGYINTKGRAKIL 384
 QY 371 GGGIAADRGYFIQPTVSGDGGMTIAKEETFGVMQILAKFTIEEVGGNNSTYGLAA 430
 DB 385 GGGIAADRGYFIQPTVSGDGGMTIAKEETFGVMQILAKFTIEEVGGNNSTYGLAA 444
 QY 431 AVFTKDLKANYLSQALQAGTVMVNCYDVFCAQSPFGGKMSGSGRELGEYGLQAYTEVK 490
 DB 445 AVFTKDLKANYLSQALQAGTVMVNCYDVFCAQSPFGGKMSGSGRELGEYGLQAYTEVK 504
 QY 491 TVTVKVPQKN 500
 DB 505 TVTVKVPQKN 514

RESULT 9

US-09-791-537-71773
 : Sequence 11/14, Application US/09791537
 : GENPAT. INFORMATION:
 : APPLICANT: Biomimix, Inc.
 : APPLICANT: Debe, Derek
 : APPLICANT: Danzer, Joseph
 : TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 : TITLE OF INVENTION: METHODS OF USE THEREOF
 : FILE REFERENCE: 261/210
 : CURRENT APPLICATION NUMBER: US/09/791,537
 : CURRENT FILING DATE: 2001-02-22
 : NUMBER OF SEQ ID NOS: 153055
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 71773
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : US-09-791-537-71773

Query Match 97.9%, Score 2569, DB 21, Length 517,
 Best Local Similarity 98.9%, pred. No. 1.6e-256;
 Matches 490, Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNOQPEVFCNOIFINNEWHDAVSPKTFPTVNPSTGEVTCQVAEGRKENV 60
 DB 17 LSAAATQAVPAPNOQPEVFCNOIFINNEWHDAVSPKTFPTVNPSTGEVTCQVAEGRKENV 76
 QY 61 DKAFEGHGAATGSGSPWPMIAASHGPIINPLADLIEPDTYLAALFTLNGKPYVVISLV 120
 DB 77 DKAFGAAPAPQAGSPWPMIAASHGPIINPLADLIEPDTYLAALFTLNGKPYVVISLV 136
 QY 121 VQAGWMLKGLRYAGWAHKKYHKTPTPTGTFEFSYTPRHPVSVVGGTTPWNEPLMLKAWK 180
 DB 137 VQAGWMLKGLRYAGWAHKKYHKTPTPTGTFEFSYTPRHPVSVVGGTTPWNEPLMLKAWK 196
 QY 181 GPALATGNNVVMKVAEFTPTALYVANIIEKAGPPGVVNIIVPGFPTAGAAATASHETV 240
 DB 197 GPALATGNNVVMKVAEFTPTALYVANIIEKAGPPGVVNIIVPGFPTAGAAATASHETV 256
 QY 241 KVAFGSGTEFGVTCVAAGSSNIPKVTLEFGSKSFNINSLALMIMWAVEKAHFALEFNG 300
 DB 257 KVAFGSGTEFGVTCVAAGSSNIPKVTLEFGSKSFNINSLALMIMWAVEKAHFALEFNG 316
 QY 301 QCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLE 360
 DB 317 QCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLE 376
 QY 361 QCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLE 420
 DB 377 QCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLEFNGQCCAGSPLE 436
 QY 431 ANNSYGLAAAVFEGKLLKANYLSQALQAGTVMVNCYDVFCAQSPFGGKMSGSGRELGE 480
 DB 447 ANNSYGLAAAVFEGKLLKANYLSQALQAGTVMVNCYDVFCAQSPFGGKMSGSGRELGE 496
 QY 491 YGLQAYTEVKTVTVKVPQKN 500
 DB 497 YGLQAYTEVKTVTVKVPQKN 516

RESULT 10

US-50-389-987-56
 : Sequence 56, Application US/60389987
 : GENERAL INFORMATION:
 : APPLICANT: Ghosh, Soumitra S.
 : APPLICANT: Faby, Eoin D.
 : APPLICANT: Zhang, Bing
 : APPLICANT: Gibson, Bradford W.
 : APPLICANT: Taylor, Steven W.
 : APPLICANT: Glenn, Gary M.
 : APPLICANT: Warnock, Dale E.
 : TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

[illegible]

RESULT 13

```

RESOLUT 13
US-09-791-537-64443
: Sequence 64443. Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Biomomix, Inc.
: APPLICANT: Debe, Detek
: APPLICANT: Danzef, Joseph
: TITLE OF INVENTION: THREE DIMENSIONAL STRUTS
: TITLE OF INVENTION: METHODS OF USE THEREOF
: FILE REFERENCE: 261/210
: CURRENT APPLICATION NUMBER: US/09/791,537
: CURRENT FILING DATE: 2001-02-22
: NUMBER OF SEQ IN NOS: 14005
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 64443
: LENGTH: 500
: TYPE: prt
: ORGANISM: Mesocricetus auratus
US-09-791-537-64443

```

Query Match	95.4%	Score 2503	DB 21	Length 500
Best Local Similarity	95.0%	Pred. No. 11e-249		
Matches 474	Conservative 11	Mismatches 14	Indels 0	Gaps 0
QY	2	SAAATQAPAPNQPEVFCNCIFINNEHDAVSKTPTPVNPSGTGCVITDVAEGSKEUVD	61	
DB	1	SAAATSAVAPAPNQPEVFCNCIFINNEHDAVSKTPTPVNPSGTGCVITDVAEGSKEUVD	60	
QY	62	KAPGKRPAPPTGSPWPMASHSCPLNPLADLPDPTYLAALETLNGRPYVTSYLV	121	
DB	61	KAVKAAAPAFSGSPWSPMDASGRGLNPLADLPDPTYLAALETLNGRPYVTSYLV	120	
QY	122	LDLDMWLKTLPEYATGALKYHKKTIPTGSGKPSYTRHPGVSVGQILPWNPTGLMKAWKLK	181	
DB	121	LDLDMWLKTLRYATGALKYHKKTIPTGSGKPSYTRHPGVSVGQILPWNPTGLMKAWKLK	180	
QY	182	PALATCNVVMKVAEQTLTALYVANIKEASFPFQVNNIVGPGPTAGCAALASHEEVLR	241	
DB	181	PALATGNVVMKVAEQTLTALYVANIKEASFPFQVNNIVGPGPTAGCAALASHEEVLR	240	

[illegible]

Query Match	95.24	Score 2509	DB 21	Length 519
Best Local Similarity	94.8%	Prod. No. 2.30e+249		
Matches 474	Conservative 12	Mismatches 14	Indels 0	Gaps 0
QY	1	MSAATLAVAPAPNQLPEVFNQIFINNEHDAVSPTKPIIVNPSFVFIQGVAFSDKEDV	60	
DB	19	LSAATASVAPAPNQQPVFCNQIFINNEHDAVSKKTPPTVNPSTGVCICVQAVGKEDV	78	
QY	61	IKAFEPSPFAPQIGSPSPRRMTASHSPINLPIATLTPPERTYLAALFTLDGKQPVYSYL	120	
DB	79	DKAVKAAQAAPQLGSPWPPNDASDPGLLYPLADLIERPTYLAALFTLDGKQPVYSYL	138	
QY	121	VFLDMVLKTLPPYAAQWAKTHSKLPIIGTIFSTSYTHHEPVVWVQQIIPWNPPTLMQAWKL	180	
DB	139	VOLDMLVKCLRYAGWADKYHGKTIPIIDGDTFSYTHHEPVGVCGQIIPWNPPTLMQAWKL	198	
QY	181	SPALATSNVVVMKVAEYTPLTALYVANLKEAGFPFPGVYVNIVPFSGPTAGAAIASHEVDV	240	
DB	199	SPALATSNVVVMKVAEYTPLTALYVANLKEAGFPFPGVYVNIVPFSGPTAGAAIASHEVDV	258	
QY	241	KVAFPGSTFEGVTVQVAGSSNLKPVTLPIQCKSPNIINSDADMIWAVFQAHFALPNNQC	300	
DB	259	KVAFPGSTFEGVHLTVQVAGSSNIPVTLPIQCKSPNIINSDADMIWAVFQAHFALPNNQC	318	
QY	301	QVQVAGSPTFPQEDIDYDFEPPVPSVAVAKSPVVIINPEPSKTPQSPQVHVFQCKKILGYLNF	360	
DB	319	QVQVAGSPTFPQELIVYDFEPPVPSVAVAKSPVVIINPEPSKTPQSPQVHVFQCKKILGYLNF	378	
QY	361	QKQSAKLLQGGGIAADGGYFIQPIVFGWQWQIMTIAKKEEIPSPVMQILKTKTIPFVWGP	420	
DB	379	QKQSAKLLQGGGIAADGGYFIQPIVFGWQWQIMTIAKKEEIPSPVMQILKTKTIPFVWGP	438	

QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVWVNCYDVFAGSPGGYKMSGSGRELGE 480
DB 439 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVWVNCYDVFAGSPGGYKMSGSGRELGE 498
QY 481 YGLQAYTEVKTIVTKVPQKN 500
DB 499 YGLQAYTEVKTIVTKVPQKN 518

RESULT 15

US-09-791-537-130904
; Sequence 130904, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; NUMBER OF SEQ ID NOS: 153095
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 130904
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-791-537-130904

Query Match 94.5%; Score 2480; DB 21; Length 519;
Best Local Similarity 94.2%; Pred. No. 2.8e-247;
Matches 471; Conservative 13; Mismatches 16; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGCVICQVAGCKEDV 60
DB 19 LSAAATSAVPAPNHQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGCVICQVAGCKEDV 78
QY 61 DKAFEGPAPFQAGSPWPPMDASHGRLLNRLADLIERDPTYLAALETLDNCKPYVSYL 120
DB 79 DKAVKAARAFAQJGSPWRMDASDRGLLYRLADLIERDPTYLAALETLDNCKPYVSYL 138
QY 121 VLDLWVLKCLRYAGWADKYHGKTIPIDGDFESYTRHEPVGCGQIIPWNEPFLMQAWKL 180
DB 139 VLDLWVLKCLRYAGWADKYHGKTIPIDGDFESYTRHEPVGCGQIIPWNEPFLMQAWKL 198
QY 181 GPALATCNVVMKVAQTPUTALYVANLIKEAGFPVGVNIVPGFGPTAGAAIASHEVDV 240
DB 199 GPALATCNVVMKVAQTPUTALYVANLIKEAGFPVGVNIVPGFGPTAGAAIASHEVDV 258
QY 241 KVAFTGSTEGRVIGVAAAGSSNIKRVITLGGKSPNIMSDADMDWAVEQAHFALFTNQ 300
DB 259 KVAFTGSTEGRVIGVAAAGSSNIKRVITLGGKSPNIMSDADMDWAVEQAHFALFTNQ 318
QY 301 QCCAGSRFTVQEDIDYDFEVSVAPAKSPVGNPDSKTEGPGVDFIQPKILGYINT 360
DB 319 QCCAGSRFTVQEDIDYDFEVSVAPAKSPVGNPDSKTEGPGVDFIQPKILGYINT 378
QY 361 GQEGAKLLCGGFIADRGYFIQPTVFGVQDMTIAKEEIPGPMQILKEKTIIEVVGR 420
DB 379 GQEGAKLLCGGGAADRGYFIQPTVFGVQDMTIAKEEIPGPMQILKEKTIIEVVGR 438
QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVWVNCYDVFAGSPGGYKMSGSGRELGE 480
DB 439 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVWVNCYDVFAGSPGGYKMSGSGRELGE 498
QY 481 YGLQAYTEVKTIVTKVPQKN 500
DB 499 YGLQAYTEVKTIVTKVPQKN 518

Search completed: June 24, 2003, 10:40:38
Job time : 183.017 secs

GenCode version 5.1.6
Copyright (c) 1995 - 2003 Compugen Inc

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:35 ; Search time 54.1289 Seconds
(without alignments)
2375 712 Million cell updates/ser

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAATQAVPAPNQPEVFC YGIGAYTEVKTIVTKVQKN 500

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1171708 seqs, 457144465 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PT_NTW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	2625	100.0	500	5	US-09-830-751-4	Sequence 4, Appl
2	2572	98.0	517	6	US-10-219-051B-6971	Sequence 6, Appl
3	2372	98.0	517	6	US-10-219-051B-6975	Sequence 6, Appl
4	2572	98.0	520	5	US-09-949-016-10653	Sequence 10653, A
5	2546.5	97.0	514	7	US-60-453-135-14871	Sequence 14871, A
6	2546.5	97.0	514	7	US-60-453-050-14871	Sequence 14871, A
7	2546.5	97.0	514	7	US-60-466-412-14871	Sequence 14871, A
8	2500	95.2	514	6	US-10-219-051B-6969	Sequence 6969, Ap
9	2500	95.2	514	6	US-10-219-051B-6973	Sequence 6973, Ap
10	1794	68.2	518	6	US-10-144-779-472	Sequence 472, Ap
11	1791	68.2	532	5	US-60-443-615-6207	Sequence 6207, Ap
12	1790	68.2	518	5	US-09-724-676-64510	Sequence 64510, A
13	1790	68.2	518	5	US-09-724-676A-64510	Sequence 64510, A
14	1774	67.6	501	5	US-09-724-676-64517	Sequence 64517, A
15	1774	67.6	501	5	US-09-724-676A-64517	Sequence 64517, A
16	1769	67.4	500	6	US-10-219-051B-12834	Sequence 12834, A
17	1769	67.4	500	6	US-10-219-051B-12838	Sequence 12838, A
18	1756.5	66.9	516	5	US-09-724-676-64518	Sequence 64518, A
19	1756.5	66.9	516	5	US-09-724-676A-64518	Sequence 64518, A
20	1756.5	66.9	516	5	US-09-724-676A-64518	Sequence 64518, A
21	1756.5	66.9	516	5	US-09-724-676A-64518	Sequence 64518, A
22	1718	65.4	512	1	PTC US02-32371-2	Sequence 2, Appl
23	1718	65.4	512	1	US-10-264-518-2	Sequence 2, Appl
24	1716	65.4	501	6	US-10-316-253-18	Sequence 18, Appl
25	1711	65.2	500	6	US-10-219-051B-12832	Sequence 12832, A
26	1711	65.2	500	6	US-10-219-051B-12836	Sequence 12836, A

ALIGNMENTS

RESULT 1

US-09-830-751-4

: Sequence 4, Application US/09830751

: GENERAL INFORMATION:

: APPLICANT: Suthers, Patrick F.

: APPLICANT: Cameron, Douglas C.

: TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant

: TITLE OF INVENTION: Organisms

: FILE REFERENCE: 99-0295 96617

: CURRENT APPLICANT NUMBER: US/09/830,751

: CURRENT FILING DATE: 2000-08-30

: PRIOR APPLICANT NUMBER: 65/151,440

: PRIOR FILING DATE: 1999-08-30

: PRIOR APPLICATION NUMBER: PCT/US00/23878

: PRIOR FILING DATE: 2000-08-30

: NIMRP OF SEQ ID NOS: 23

: SOFTWARE: Patent In Ver. 2.1

: SEQ ID NO: 4

: LENGTH: 500

: TYPE: PRT

: ORGANISM: Homo sapiens

US-09-830-751-4

Query Match	Best Local Similarity	Score	DB 5;	Length	500;
Matches	500;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;		
QY	1	MSAATQAVPAPNQPEVFCNQIFINNEHDAVSRRKTFPTVNPSTGEVLCQVAESIKEDV	60		
DB	1	MSAATQAVPAPNQPEVFCNQIFINNEHDAVSRRKTFPTVNPSTGEVLCQVAESIKEDV	60		
QY	61	TKARPGRRGAPQIGSPRRMDSHSPILNPLADI.TERTPTYLAALETLNKPKPVYSTL	120		
DB	61	TKARPGRRGAPQIGSPRRMDSHSPILNPLADI.TERTPTYLAALETLNKPKPVYSTL	120		
QY	121	VLLQMWLKEG.PYYACWAFYHGTPTTGGFFSYTHPEFVGVVGGQ.LPNWPLLMGAWKL	180		
DB	121	VLLQMWLKEG.PYYACWAFYHGTPTTGGFFSYTHPEFVGVVGGQ.LPNWPLLMGAWKL	180		
QY	241	KVAFQSGTEIFGPIVQVAAGSSNLKPKPTLELCRSHNLIMSQAQMWAVFQAPAL.FFNQ	400		
DB	241	KVAFQSGTEIFGPIVQVAAGSSNLKPKPTLELCRSHNLIMSQAQMWAVFQAPAL.FFNQ	400		
QY	301	QCCAGSKTFQEDLY.EEYVVSVAFAKSHVV.NPFLSKTEGJLVVLELQEKKKLLSYLNL	360		
DB	301	QCCAGSKTFQEDLY.EEYVVSVAFAKSHVV.NPFLSKTEGJLVVLELQEKKKLLSYLNL	360		

Db 301 QCCACGSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTCUQPOVDETQPKKILGYINT 360
 QY 361 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 361 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 480
 Db 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 480
 QY 481 YGLQAYTEVKTIVKVPQKN 500
 Db 481 YGLQAYTEVKTIVKVPQKN 500

RESULT 2

US-10-219-051B-6971
 : Sequence 6971, Application US/10219051B
 : GENERAL INFORMATION:
 : APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 : APPLICANT: Hospital / Bayer AG
 : TITLE OF INVENTION: Nucleotide sequences involved in pain
 : FILE REFERENCE: Lea 35693 Foreign Countries
 : CURRENT APPLICATION NUMBER: US/10/219,051B
 : PRIOR FILING DATE: 2003-05-09
 : PRIOR APPLICATION NUMBER: US 60/312,147
 : PRIOR FILING DATE: 2001-08-14
 : PRIOR APPLICATION NUMBER: US 60/346,382
 : PRIOR FILING DATE: 2001-11-01
 : PRIOR APPLICATION NUMBER: US 60/333,347
 : PRIOR FILING DATE: 2001-11-26
 : NUMBER OF SEQ ID NOS: 14715
 : SOFTWARE: Perl script
 : SEQ ID NO 6971
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : PUBLICATION INFORMATION:
 : DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 : DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6971

Query Match: 98.0%; Score 2572; DB 6; Length 517;
 Best Local Similarity 98.2%; Pctd. No. 4,96,224;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPFTVNPSTGEVICOVAEGKEDV 60
 Db 17 LSAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPFTVNPSTGEVICOVAEGKEDV 76
 QY 61 DKAREGRGAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETLDNGKPYVISYL 120
 Db 77 DKAVKAAARAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETLDNGKPYVISYL 136
 QY 121 VLDLMDVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGCGOIIPWNEPLLMOAWKL 180
 Db 137 VLDLMDVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGCGOIIPWNEPLLMOAWKL 196
 QY 181 GPALATGNVVMKVAEOTPLTALYVANLKEAGFPVGVNIVPGPGTAGAAIASHEIDVD 240
 Db 197 GPALATGNVVMKVAEOTPLTALYVANLKEAGFPVGVNIVPGPGTAGAAIASHEIDVD 256
 QY 241 KVAFTGSTELGRVIOVAAGSSNLKRVTTLELGKSPNIIIMSADMDWAVEQAHFALFNQ 300
 Db 257 KVAFTGSTELGRVIOVAAGSSNLKRVTTLELGKSPNIIIMSADMDWAVEQAHFALFNQ 316
 QY 301 QCCACGSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTCUQPOVDETQPKKILGYINT 360
 Db 317 QCCACGSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTCUQPOVDETQPKKILGYINT 376
 QY 361 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 377 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 436

QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 480
 Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 496
 QY 481 YGLQAYTEVKTIVKVPQKN 500
 Db 497 YGLQAYTEVKTIVKVPQKN 516

RESULT 3

US-10-219-051B-6975
 : Sequence 6975, Application US/10219051B
 : GENERAL INFORMATION:
 : APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 : APPLICANT: Hospital / Bayer AG
 : TITLE OF INVENTION: Nucleotide sequences involved in pain
 : FILE REFERENCE: Lea 35693 Foreign Countries
 : CURRENT APPLICATION NUMBER: US/10/219,051B
 : PRIOR FILING DATE: 2003-05-09
 : PRIOR APPLICATION NUMBER: US 60/412,147
 : PRIOR FILING DATE: 2001-08-14
 : PRIOR APPLICATION NUMBER: US 60/446,382
 : PRIOR FILING DATE: 2001-11-01
 : PRIOR APPLICATION NUMBER: US 60/433,447
 : PRIOR FILING DATE: 2001-11-26
 : NUMBER OF SEQ ID NOS: 14715
 : SOFTWARE: Perl script
 : SEQ ID NO 6975
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : PUBLICATION INFORMATION:
 : DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 : DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6975

Query Match: 98.0%; Score 2572; DB 6; Length 517;
 Best Local Similarity 98.2%; Pctd. No. 4,96,224;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPFTVNPSTGEVICOVAEGKEDV 60
 Db 17 LSAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPFTVNPSTGEVICOVAEGKEDV 76
 QY 61 DKAREGRGAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETLDNGKPYVISYL 120
 Db 77 DKAVKAAARAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETLDNGKPYVISYL 136
 QY 121 VLDLMDVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGCGOIIPWNEPLLMOAWKL 180
 Db 137 VLDLMDVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGCGOIIPWNEPLLMOAWKL 196
 QY 181 GPALATGNVVMKVAEOTPLTALYVANLKEAGFPVGVNIVPGPGTAGAAIASHEIDVD 240
 Db 197 GPALATGNVVMKVAEOTPLTALYVANLKEAGFPVGVNIVPGPGTAGAAIASHEIDVD 256
 QY 241 KVAFTGSTELGRVIOVAAGSSNLKRVTTLELGKSPNIIIMSADMDWAVEQAHFALFNQ 300
 Db 257 KVAFTGSTELGRVIOVAAGSSNLKRVTTLELGKSPNIIIMSADMDWAVEQAHFALFNQ 316
 QY 301 QCCACGSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTCUQPOVDETQPKKILGYINT 360
 Db 317 QCCACGSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTCUQPOVDETQPKKILGYINT 376
 QY 361 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 377 GKQGRKLLGGGTAAGRGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 436
 QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 480
 Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVFAGSPGGYKMSGSGREIGE 496

```

QY 481 YGLOAYTEVKTIVTKVPOKN 500
Db 497 YGLOAYTEVKTIVTKVPOKN 516

RESULT 4
US-09-949-016-10653
; Sequence 10653, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10653
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10653

```

```

Query Match 98.08, Score 2572, DB 5, Length 520,
Best Local Similarity 98.24, Prod No. 4,900,224,
Matches 491, Conservative 2, Mismatches 7, Indels 0, Gaps 0.

QY 1 MSAAATQAVPAPNQQPEVFCNQIFINNEHDAVSRKTPTVNPSTGEVICOVAEGDKEDV 60
Db 20 LSAAATQAVPAPNQQPEVFCNQIFINNEHDAVSRKTPTVNPSTGEVICOVAEGDKEDV 79
QY 61 DKAPEPGGAPFQIGSPWPRMDASHGRLNPLADLIERDPTYLAAETLDNGKPVVISYL 120
Db 80 DKAVKAAPAAAFQIGSPWPRMDASHGRLNPLADLIERDPTYLAAETLDNGKPVVISYL 139
QY 121 VLDLMLKCLRYTAGWADKYHGKTIPIDGDFSYTHHEPVGVCGQIIPWNPFLMQAKL 180
Db 140 VLDLMLKCLRYTAGWADKYHGKTIPIDGDFSYTHHEPVGVCGQIIPWNPFLMQAKL 199
QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 240
Db 200 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 259
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLDELGKSPNII MSDADMDWAVEQAHFALFFNOG 300
Db 260 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLDELGKSPNII MSDADMDWAVEQAHFALFFNOG 319
QY 301 QCCAGSPTFVQELIYDEFEVVSVAKSRVGVNPNFDSKTEGQPVDETFQKKILGYINT 360
Db 320 QCCAGSPTFVQELIYDEFEVVSVAKSRVGVNPNFDSKTEGQPVDETFQKKILGYINT 379
QY 361 GKUGAKLLCGGGAADRGYFIQPIVFGDVGQMTIAKBEIFGPMQILKFKTIEVVVGR 420
Db 380 GKUGAKLLCGGGAADRGYFIQPIVFGDVGQMTIAKBEIFGPMQILKFKTIEVVVGR 449
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWVNCYDVEFGAGSPFGYKMSGSGRELGE 480
Db 440 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWVNCYDVEFGAGSPFGYKMSGSGRELGE 499
QY 481 YGLOAYTEVKTIVTKVPOKN 500
Db 500 YGLOAYTEVKTIVTKVPOKN 519

RESULT 5
US-60-453-135-14871
; Sequence 14871, Application US/60453135

```

```

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOBUOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; PRIORITY FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-14871

Query Match 97.08, Score 2546.5, DB 7, Length 514,
Best Local Similarity 97.68, Prod No. 9,900,222,
Matches 488, Conservative 2, Mismatches 7, Indels 3, Gaps 1.

QY 1 MSAAATQAVPAPNQQPEVFCNQIFINNEHDAVSRKTPTVNPSTGEVICOVAEGDKEDV 60
Db 17 LSAAATQAVPAPNQQPEVFCNQIFINNEHDAVSRKTPTVNPSTGEVICOVAEGDKEDV 76
QY 61 DKAPEPGGAPFQIGSPWPRMDASHGRLNPLADLIERDPTYLAAETLDNGKPVVISYL 120
Db 77 DKAVKAAPAAAFQIGSPWPRMDASHGRLNPLADLIERDPTYLAAETLDNGKPVVISYL 136
QY 121 VLDLMLKCLRYTAGWADKYHGKTIPIDGDFSYTHHEPVGVCGQIIPWNPFLMQAKL 180
Db 137 VLDLMLKCLRYTAGWADKYHGKTIPIDGDFSYTHHEPVGVCGQIIPWNPFLMQAKL 195
QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 240
Db 197 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLDELGKSPNII MSEALMIWAVEQAHFALFFNOG 300
Db 257 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLDELGKSPNII MSEALMIWAVEQAHFALFFNOG 316
QY 301 QCCAGSPTFVQELIYDEFEVVSVAKSRVGVNPNFDSKTEGQPVDETFQKKILGYINT 360
Db 317 QCCAGSPTFVQELIYDEFEVVSVAKSRVGVNPNFDSKTEGQPVDETFQKKILGYINT 376
QY 361 GKUGAKLLCGGGAADRGYFIQPIVFGDVGQMTIAKBEIFGPMQILKFKTIEVVVGR 420
Db 377 GKUGAKLLCGGGAADRGYFIQPIVFGDVGQMTIAKBEIFGPMQILKFKTIEVVVGR 436
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWVNCYDVEFGAGSPFGYKMSGSGRELGE 480
Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWVNCYDVEFGAGSPFGYKMSGSGRELGE 496
QY 481 YGLOAYTEVKTIVTKVPOKN 500
Db 497 YGLOAYTEVKTIVTKVPOKN 513

RESULT 6
US-60-453-050-14871
; Sequence 14871, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: SIENOSTIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; PRIORITY FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT

```

ORGANISM: Homo sapiens
US-60-466-412-14871

Query Match 97.0%; Score 2546.5; DB 7; Length 514;
Best Local Similarity 97.6%; Pred. No. 9.9e-222;
Matches 488; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 60
DQ :|||||
DB 17 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 76
QY 61 DKAREGRPCAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 120
DQ :|||||
DB 77 DKAVKAARAAAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 136
QY 121 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 180
DQ :|||||
DB 137 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 196
QY 181 GPALATGNVVMKVAEQTPLTALYVANLIKEAGFPVGVVNIIPGEGPTAGAAIASHEDVD 240
DQ :|||||
DB 197 GPALATGNVVMKVAEQTPLTALYVANLIKEAGFPVGVVNIIPGEGPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLELGGKSPNII MSDADMDWAVEQAHFALFFNQ 300
DQ :|||||
DB 257 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLELGGKSPNII MSDADMDWAVEQAHFALFFNQ 316
QY 301 OCCAGSRTFVQDIDYDEFVRSVARAKSRVGNPFDSKTEGQPOVDETQPKKILGYINT 360
DQ :|||||
DB 317 OCCAGSRTFVQDIDYDEFVRSVARAKSRVGNPFDSKTEGQPOVDETQPKKILGYINT 376
QY 361 GKQEGAKLLCGGGIAADRGYFTQPTVFGDVQDQMTIAKEEIRGPVMQILKFKTIEEVVGR 420
DQ :|||||
DB 377 GKQEGAKLLCGGGIAADRGYFTQPTVFGDVQDQMTIAKEEIRGPVMQILKFKTIEEVVGR 436
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVVWNYIVFSAQSPFGYKMGSGSRELGE 480
DQ :|||||
DB 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVVWNYIVFSAQSPFGYKMGSGSRELGE 496
QY 481 YGLQAYTEVKTIVTKVPQKN 500
DQ :|||||
DB 497 YGLQAYTE---VTVKVPQKN 513

```

RESULT 7

```

US-60-466-412-14871
: Sequence 14871, Application US/60466412
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001466
: CURRENT APPLICATION NUMBER: US/60/466,412
: NUMBER OF SEQ ID NOS: 429241
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14871
: LENGTH: 514
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-466-412-14871

```

Query Match 97.0%; Score 2546.5; DB 7; Length 514;
Best Local Similarity 97.6%; Pred. No. 9.9e-222;
Matches 488; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 60
DQ :|||||
DB 17 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 76
QY 61 DKAREGRPCAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 120
DQ :|||||

```

```

DB 77 DKAVKAARAAAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 146
QY 121 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 180
DQ :|||||
DB 137 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 196
QY 181 GPALATGNVVMKVAEQTPLTALYVANLIKEAGFPVGVVNIIPGEGPTAGAAIASHEDVD 240
DQ :|||||
DB 197 GPALATGNVVMKVAEQTPLTALYVANLIKEAGFPVGVVNIIPGEGPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLELGGKSPNII MSDADMDWAVEQAHFALFFNQ 300
DQ :|||||
DB 257 KVAFTGSTEIGRVIOVAAGSSNLKRVTLLELGGKSPNII MSDADMDWAVEQAHFALFFNQ 316
QY 301 OCCAGSRTFVQDIDYDEFVRSVARAKSRVGNPFDSKTEGQPOVDETQPKKILGYINT 360
DQ :|||||
DB 317 OCCAGSRTFVQDIDYDEFVRSVARAKSRVGNPFDSKTEGQPOVDETQPKKILGYINT 376
QY 361 GKQEGAKLLCGGGIAADRGYFTQPTVFGDVQDQMTIAKEEIRGPVMQILKFKTIEEVVGR 420
DQ :|||||
DB 377 GKQEGAKLLCGGGIAADRGYFTQPTVFGDVQDQMTIAKEEIRGPVMQILKFKTIEEVVGR 436
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVVWNYIVFSAQSPFGYKMGSGSRELGE 480
DQ :|||||
DB 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVVWNYIVFSAQSPFGYKMGSGSRELGE 496
QY 481 YGLQAYTEVKTIVTKVPQKN 500
DQ :|||||
DB 497 YGLQAYTE---VTVKVPQKN 513

```

RESULT 8

```

US-10-219-051B-6969
: Sequence 6969, Application US/10219051B
: GENERAL INFORMATION:
: APPLICANT: The General Hospital Corporation, doing business as Massachusetts General
: TITLE OF INVENTION: Nucleotide sequences involved in pain
: FILE REFERENCE: LEA 45694 Foreign Countries
: CURRENT APPLICATION NUMBER: US/10/219,051B
: CURRENT FILING DATE: 2003-05-09
: PRIOR APPLICATION NUMBER: US 60/412,147
: PRIOR FILING DATE: 2001-08-14
: PRIOR APPLICATION NUMBER: US 60/446,482
: PRIOR FILING DATE: 2001-11-01
: PRIOR APPLICATION NUMBER: US 60/443,447
: PRIOR FILING DATE: 2001-11-26
: NUMBER OF SEQ ID NOS: 14715
: SOFTWARE: Perl script
: SEQ ID NO 6969
: LENGTH: 519
: TYPE: PRT
: ORGANISM: Rattus norvegicus
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
: DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-6969

```

Query Match 95.2%; Score 2500; DB 6; Length 519;
Best Local Similarity 94.8%; Pred. No. 1.6e-217;
Matches 474; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 60
DQ :|||||
DB 19 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKIFPIVNI-STG:EVICUVAEGDKEDV 78
QY 61 DKAREGRPCAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 120
DQ :|||||
DB 79 DKAVKAARAAAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPYVSYL 148
QY 121 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 180
DQ :|||||
DB 149 VLDMLVKCLRYAYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCGQIIPWNPFLLMQAWKL 198

```


Db 441 SDFGLVAAVFTNDINKALMVSSAMQAGTVMWNCYNALNAQSPFGGFKMSGNCREMGEGCL 500

QY 484 QAYTEVKTVTVKIPQKN 500

Db 501 REYSEVKTVTVKIPQKN 517

RESULT 11

US-09-949-016-9207

Sequence 9207, Application US/09949016

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

CURRENT APPLICATION NUMBER: 60/001307

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9207

LENGTH: 532

TYPE: PRT

ORGANISM: Human

US-09-949-016-9207

Query Match 68.2%, Score 1791, DB 5, Length 532;

Best Local Similarity 66.2%, Pred. No. 3, 1e-153;

Matches 329; Conservative 75; Mismatches 93; Indels 0; Gaps 0;

QY 4 AATQAVPAPNQOPEVFCNQIFINNEWHDAVSARKTEFTVNPSTGEGVICOVAGCKEDVDVKA 63

Db 35 ASUHLPSPTNLEIKYKIFINNEWQNSGKVFVNPAIGEVCEVQADKADIDKA 94

QY 64 REGPGAFQSGWRRMDASHGRLLNRLADLIERDRTYLAALFTLDNCKPVPVSYLDL 123

Db 95 VQARLAFSLGSGVWRMDASERGLLDLADLVERDRAVLATMESLNGSKPFLQAFYDL 154

QY 124 DMVLKCLRYAGWADYHCKTIPIDGDFSYTRHEPVGCGGIIIPWNPFLLMQAWKLGPA 183

Db 155 QGVKTFRYAGWADKHGMITIPVDGDFYTRHEPVGCGGIIIPWNPFLLMQAWKLGPA 214

QY 184 LATGNVVMKVAEOTPLTALYVANLIKAGPPGPPVNVIPGPGTAGAAIASHEDVDKVA 243

Db 215 LCCGNTVWIKPAEQIFPLSALYKALIKAGPPGPPVNVIPGPGTAGAAIASHIDKIA 274

QY 244 FTGSTEIGRVIOVAGGSSNLKRVTLIELGCKSPNIIIMSDADMWAVEQAHFALFTNQGQC 303

Db 275 FTGSTEVGKLIQEAAGRSNLKRVTLIELGCKSPNIIIFADLDLYAVEQAHQGVFFNQGC 334

QY 304 CAGSRFTVQEDIIYDFVVRVSARAKSRVGVNPFDSKTFQGGQVDETQPKKILGINTGK 363

Db 335 TAGSRIFVFSIEYEFVVRVSARAKSRVGVNPFDSKTFQGGQVDETQPKKILGINTGK 394

QY 364 EGAKLCCGGIADRGYFIQPTVFGDVQGMIIAKEELFGVMQILKFKTEEEVVGGRANN 423

Db 395 EGAKLECGKGLGRKGFIEPTVFSNVTDMMRIAKEELFGVQVLEIRKTEDEVIERANN 454

QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFGAQSIPFGYKMSGSGRELGEYGL 483

Db 455 SDFGLVAAVFTNDINKALMVSSAMQAGTVMWNCYNALNAQSPFGGFKMSGNGREMGEGCL 514

QY 484 QAYTEVKTVTVKIPQKN 500

Db 515 REYSEVKTVTVKIPQKN 531

RESULT 12

US-09-724-676-64510

Sequence 64510, Application US/09724676

GENERAL INFORMATION:

APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing

FILE REFERENCE: 129181.4 Compugen

CURRENT APPLICATION NUMBER: US/09/724,676

CURRENT FILING DATE: 2000-11-28

NUMBER OF SEQ ID NOS: 97222

SOFTWARE: PatentIn version 3.2

SEQ ID NO 64510

LENGTH: 518

TYPE: PRT

ORGANISM: Homo sapiens

US-09-724-676-64510

Query Match 68.2%, Score 1790, DB 5, Length 518;

Best Local Similarity 66.0%, Pred. No. 3, 6e-153;

Matches 328; Conservative 76; Mismatches 93; Indels 0; Gaps 0;

QY 4 AATQAVPAPNQOPEVFCNQIFINNEWHDAVSARKTEFTVNPSTGEGVICOVAGCKEDVDVKA 63

Db 21 ASUHLPSPTNLEIKYKIFINNEWQNSGKVFVNPAIGEVCEVQADKADIDKA 80

QY 64 REGPGAFQSGWRRMDASHGRLLNRLADLIERDRTYLAALFTLDNCKPVPVSYLDL 123

Db 81 VQARLAFSLGSGVWRMDASERGLLDLADLVERDRAVLATMESLNGSKPFLQAFYDL 140

QY 124 DMVLKCLRYAGWADYHCKTIPIDGDFSYTRHEPVGCGGIIIPWNPFLLMQAWKLGPA 183

Db 141 QGVKTFRYAGWADKHGMITIPVDGDFYTRHEPVGCGGIIIPWNPFLLMQAWKLGPA 200

QY 184 LATGNVVMKVAEOTPLTALYVANLIKAGPPGPPVNVIPGPGTAGAAIASHEDVDKVA 243

Db 201 LCCGNTVWIKPAEQIFPLSALYKALIKAGPPGPPVNVIPGPGTAGAAIASHIDKIA 260

QY 244 FTGSTEIGRVIOVAGGSSNLKRVTLIELGCKSPNIIIMSDADMWAVEQAHFALFTNQGQC 303

Db 261 FTGSTEVGKLIQEAAGRSNLKRVTLIELGCKSPNIIIFADLDLYAVEQAHQGVFFNQGC 320

QY 304 CAGSRFTVQEDIIYDFVVRVSARAKSRVGVNPFDSKTFQGGQVDETQPKKILGINTGK 363

Db 321 TAGSRIFVFSIEYEFVVRVSARAKSRVGVNPFDSKTFQGGQVDETQPKKILGINTGK 380

QY 364 EGAKLCCGGIADRGYFIQPTVFGDVQGMIIAKEELFGVMQILKFKTEEEVVGGRANN 423

Db 381 EGAKLECGKGLGRKGFIEPTVFSNVTDMMRIAKEELFGVQVLEIRKTEDEVIERANN 440

QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFGAQSIPFGYKMSGSGRELGEYGL 484

Db 441 SDFGLVAAVFTNDINKALMVSSAMQAGTVMWNCYNALNAQSPFGGFKMSGNGREMGEGCL 500

QY 484 QAYTEVKTVTVKIPQKN 500

Db 501 REYSEVKTVTVKIPQKN 517

RESULT 13

US-09-724-676A-64510

Sequence 64510, Application US/09724676A

GENERAL INFORMATION:

APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing

FILE REFERENCE: 129181.4 Compugen

CURRENT APPLICATION NUMBER: US/09/724,676A

CURRENT FILING DATE: 2000 11 28

NUMBER OF SEQ ID NOS: 97222

SOFTWARE: PatentIn version 3.2

SEQ ID NO 64510

LENGTH: 518

TYPE: PRT

ORGANISM: Homo sapiens

US-09-724-676A-64510


```
Db 361 GKKEGAKIFCGGGEWGNKGYFVQPTVFSNVTDEMRTAKEEIFGPVQIMKFKSLDDVIKR 420
QY 421 ANNSTYGLAAVFTKDLKANYLSQALQACTVWVNCYDVFAGOSPPGGYKMSGSGKEIGE 480
Db 421 ANNTFYGLSAGVETKIDKAITISSALQACTVWVNCYGVVSACCPGGFKMSGNGRELGE 480
QY 481 YGLQAYTEVKTVTVKVPKN 500
Db 481 YGFHEYTEVKTVTVKISOKN 500
```

Search completed: June 24, 2003, 10:35:21
Job time : 56.1289 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 GenCore Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:10:05 ; Search time 14.145 seconds
(without alignments)
2512.114 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 242

Sequence: 1 MSAATCAVAPAPACQPEVFC YRLQAVTEVKTIVKVPQFN 500

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_73:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2572	98.0	517	1	DEH0E2
2	2500	95.2	519	1	SS3564
3	2480	94.5	519	1	148966
4	2436	92.8	520	1	S09030
5	2434	92.7	530	1	S00364
6	1939	73.9	517	1	A40872
7	1802	64.4	504	1	S14629
8	1794	64.4	494	2	S74224
9	1774	67.6	501	1	DEH0E1
10	1763	67.2	501	2	S14752
11	1757	66.9	501	2	SC5553
12	1744	66.4	501	1	I01004
13	1743	66.4	501	2	JC4524
14	1736	66.1	500	1	S02302
15	1718	65.4	512	1	A55684
16	1716	65.4	501	1	A32616
17	1703.5	64.9	544	2	D88449
18	1587	60.5	497	2	374924
19	1543	58.8	538	2	T06483
20	1539	58.6	542	2	T02361
21	1537	58.6	549	2	T04983
22	1499	57.1	496	1	A46725
23	1498.5	57.1	514	2	D86472
24	1425.5	54.3	497	1	SC9055
25	1422.5	54.2	511	1	S31308
26	1378.5	52.5	533	2	T54216
27	1354	51.6	498	2	C83717
28	1325	50.5	495	2	H60614
29	1305.5	49.7	496	2	S43114

30	1269	48.3	495	2	S43108
31	1235.5	47.1	519	1	S67286
32	1199.5	45.7	498	1	S43184
33	1178	44.9	496	2	T50272
34	1171	44.6	520	1	S50576
35	1133.5	43.2	500	1	S60039
36	1083	41.3	506	1	S54615
37	1080	41.3	506	1	S54527
38	1035	39.4	494	2	A54624
39	1034.5	39.4	902	2	A60560
40	1021	38.9	497	2	R82481
41	1016.5	38.7	495	2	H83136
42	1001.5	38.3	503	1	S15335
43	991.5	37.8	486	2	H83131
44	990	37.7	496	2	C87638
45	989	37.7	497	1	A35994

ALIGNMENTS

RESULT 1

DEH0E2

A:aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial [validated] - hu
N:Alternate names: aldehyde dehydrogenase E2; aldehyde dehydrogenase I
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1988 #sequence revision 17-Nov-1995 #rev. change 03-Jun-2002
C:Accession: A29375, S00804, A23503, A27509, A26743, L39433, L39430
R:Hsu, L.C.; Bengel, R.F.; Yoshida, A
Genomics 2, 57-65, 1988
A:Title: Genomic structure of the human mitochondrial aldehyde dehydrogenase gene.
A:Reference number: A29375, M01D:88256153, F01D:2838413
A:Accession: A29975
A:Molecule type: DNA
A:Residues: 1-517 <HSU1>
A:Cross-references: R: M06760, NID:317837, PIR: AAA51694 1; PID:317837
R:Braun, T.; Heber, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
FEBS Lett. 233, 440, 1988
A:Reference number: S00804
A:Accession: S00804
A:Molecule type: mRNA
A:Residues: 1-6, 'APA', 10, 'F', 13, 49, 'BPA'
A:Cross-references: EMBL:X55409, NID:328605
A:Note: correction to A26743
R:Bengel, J.; Kaiser, R.; Joernvall, H.
Eur. J. Biochem. 153, 13-28, 1985
A:Title: Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, di
A:Reference number: A23503; M01D:86055846, F01D:4065146
A:Accession: A23503
A:Molecule type: protein
A:Residues: 'A', 15-517, 'HEM'
A:Note: the sequence shown is presumably that of the mature protein, however, the num
R:Braun, T.; Heber, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
Nucleic Acids Res. 15, 3179, 1987
A:Title: Isolation and sequence analysis of a full length cDNA clone coding for human
A:Reference number: A27509, M01:8714847, F01D:362530
A:Accession: A27509
A:Molecule type: mRNA
A:Residues: 1-6, 'AMFWAW', 10, 'F', 12, 'VS', 15, 'FHFGF', 21, 27-79, 'FSGPG', 86-336, 'V', 338-5
A:Cross-references: EMBL:X0109, NID:328607, FIDN:CA658290.1; FID:q28608
R:Braun, T.; Bengel, R.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
FEBS Lett. 215, 233-236, 1987
A:Title: Evidence for a signal peptide at the amino-terminal end of human mitochondrial
A:Reference number: A26743, M01:8721391, F01D:3582651
A:Accession: A26743
A:Molecule type: mRNA
A:Residues: 1-6, 'AMFWAW', 10, 'F', 12, 'VS', 15, 'FHFGF', 21, 27-79, 'FSGPG', 86-336, 'V', 338-5
A:Cross-references: EMBL:X0400, NID:924605
A:Note: this sequence is revised in reference S00804
R:Hsu, L.C.; Tai, K.; Fujiyoshi, T.; Furaihi, K.; Yoshida, A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3771-3775, 1985
A:Title: Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.
A:Reference number: L39432, M01D:85216574, F01D:2087944

A:Accession: J39432
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 119-336, V, 338-517 <HS02>
 A:Cross-references: GB:K03001, NID:q178495, P1IN:AAK54500.1; P1D:q178495
 A:Note: thirty-three tryptic peptides were also sequenced
 R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
 Alcohol 2, 103-106, 1985
 A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
 A:Reference number: J39431; MUID:85252089; PMID:4015823
 A:Accession: J39433
 A:Status: preliminary, translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 119-336, V, 338-517 <VS>
 A:Cross-references: GB:M26760, NID:q178397, P1IN:AA51634.1, P1D:q178398
 R:Agarwal, D.P.; Goede, H.W.
 Isozymes Curr Top Biol Med Res 16, 21-48, 1987
 A:Title: Human aldehyde dehydrogenase isozymes and alcohol sensitivity.
 A:Reference number: J39430; MUID:87279033; PMID:3610592
 A:Accession: J39430
 A:Status: preliminary, translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 214-215, S, 217, R, 219, 246, F, 248, 336, V, 338, 425, EVJHKGCGWESQGFHVRAJ, 426
 A:Cross-references: GB:M54841; NID:q178491; P1IN:AAK2845.1, P1D:q178492
 A:Note: the sequence is misidentified as aldehyde dehydrogenase 1
 C:Genetics:
 A:Gene: GDB:ALDH2
 A:Cross-references: GDB:119668; OMIM:100650
 A:Map position: 12q24.2-12q24.2
 A:Introns: 38/4; 74/4; 120/4; 147/2; 184/3; 227/3; 255/3; 480/1; 61/2; 416/4; 409, 2; 50
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase
 F:1-17/Domain: transit peptide (mitochondrion) #status predicted <SIC>
 F:18-517/Product: aldehyde dehydrogenase (NAD+) 2 #status experimental <MAT>
 F:75-335/Domain: aldehyde dehydrogenase homology <ALDH>
 F:211-291/Domain: NAD binding #status predicted <NAD>
 F:285/Active site: Glu #status predicted
 F:319/Active site: Cys #status experimental
 F:472/Binding site: NAD (Cys) #status predicted

Query Match 98.0%; Score 2572; DB 1; Length 517;
 Best Local Similarity 98.2%; Pred. No. 9e-191;
 Matches 491; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSAAATGAPAPNQQPEVFCNQIFINNEHDAVSRRKTPPTVNPSTGEVICQVAGDKEDV 60
 DB :|||||
 QY 17 LSAATGAPAPNQQPEVFCNQIFINNEHDAVSRRKTPPTVNPSTGEVICQVAGDKEDV 76
 DB :|||||
 QY 61 DKAREGRPGAFQCGSPWRMDASHGRGLNRLNLAJLERDITYLAALETLDNGKPYVTSYL 120
 DB :|||||
 QY 77 DKAVKAARAAPQFGSPWRMDASHGRGLNRLNLAJLERDITYLAALETLDNGKPYVTSYL 136
 DB :|||||
 QY 121 VELDMVLKCLRYAGWADYHCKTIPIDGDFSYTRHBEVVCVGGIIPWNPFLDMQAWKL 180
 DB :|||||
 QY 137 VLDLDMVLKCLRYAGWADYHCKTIPIDGDFSYTRHBEVVCVGGIIPWNPFLDMQAWKL 196
 DB :|||||
 QY 181 GPALATGNVNVKVAOTPLTALYVANLKEAGPPGVNIVPGPGTAGAATAASHEDVD 240
 DB :|||||
 QY 197 GPALATGNVNVKVAOTPLTALYVANLKEAGPPGVNIVPGPGTAGAATAASHEDVD 256
 DB :|||||
 QY 241 KVAFTGSTEIGRVIOVAASSNKPVITELGGRKSPNITMSDADIMWAVEQAHFALPTNQC 300
 DB :|||||
 QY 257 KVAFTGSTEIGRVIOVAAAGSSNKKVILELGGCKSNITMSDADIMWAVEQAHFALPTNQC 316
 DB :|||||
 QY 301 QCCACGSRFTVQEDIDYDEFWERSVARAKSRVGNVNFUSKTEGGPGVDETQPKKILGYINT 360
 DB :|||||
 QY 317 QCCACGSRFTVQEDIDYDEFWERSVARAKSRVGNVNFUSKTEGGPGVDETQPKKILGYINT 376
 DB :|||||

QY 361 GKQFGAKLLCGGGIAAURGYFTQPTVFGVQDQMTAKKEIECPVWQILKKTIEVQVR 420
 DB :|||||
 QY 377 GKQFGAKLLCGGGIAAURGYFTQPTVFGVQDQMTAKKEIECPVWQILKKTIEVQVR 436
 DB :|||||
 QY 421 ANNSTVGLAAAVFKULRWANVLSQALAG:VWVN:YGVFVAGSPFQVYKMSGSGFELGE 480
 DB :|||||
 QY 437 ANNSTVGLAAAVFKULRWANVLSQALAG:VWVN:YGVFVAGSPFQVYKMSGSGFELGE 496
 DB :|||||
 QY 481 YGLQAYTEVKTIVTKVPOKN 500
 DB :|||||
 QY 497 YGLQAYTEVKTIVTKVPOKN 516
 DB :|||||
 RESULT 2
 S03564
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Feb-1990 #sequence revision 31 Jan 1997 #text change 03 Jun 2002
 C:Accession: S03564; A:27714; S12993; S17492
 R:Farres, J.; Guan, K.L.; Weiner, H.
 Eur. J. Biochem. 180, 67-74, 1989
 A:Title: Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenase
 A:Reference number: S03564; MUID:89210865; PMID:2540003
 A:Accession: S03564
 A:Molecule type: mRNA
 A:Residues: 1-519 <FAR>
 A:Cross-references: EMBL:X14977; NID:q55604; P1IN:CAA3401.1; P1D:q55605
 A:Experimental source: Strain Sprague-Dawley; Liver
 R:Farres, J.; Guan, K.L.; Weiner, H.
 Biochem Biophys Res Commun 150, 1083-1089, 1988
 A:Title: Sequence of the signal peptide for rat liver mitochondrial aldehyde dehydrogenase
 A:Reference number: A27714; MUID:88134217; PMID:3442060
 A:Accession: A27714
 A:Molecule type: mRNA
 A:Residues: 1-29 <FAR>
 A:Cross-references: GB:M19030; NID:q202847; P1IN:AAA48719.1; P1D:q202848
 R:Diwan, J.J.; Pallwal, R.; Kaftan, E.; Bawa, R.
 FEBS Lett. 273, 215-218, 1990
 A:Title: A mitochondrial protein fraction catalyzing transport of the K(+) analog Tl(+)
 A:Reference number: S12993; MUID:91042184; PMID:1699808
 A:Accession: S12993
 A:Molecule type: protein
 A:Residues: 327-340 <DIW>
 R:Jeng, J.; Weiner, H.
 Arch. Biochem. Biophys. 289, 234-232, 1991
 A:Title: Purification and characterization of catalytically active precursor of rat 1
 A:Reference number: S17492; MUID:91378548; PMID:1898068
 A:Accession: S17492
 A:Molecule type: protein
 A:Residues: 1-19 <JEN>
 C:Genetics:
 A:Gene: nuclear
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta
 F:1-19/Domain: transit peptide (mitochondrion) #status predicted <FAR>
 F:20-519/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAT>
 F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>
 F:287,321/Active site: Glu, Cys #status predicted
 F:474/Binding site: NAD (Cys) #status predicted

Query Match 95.2%; Score 2666; DB 1; Length 519;
 Best Local Similarity 94.8%; Pred. No. 3.4e-185;
 Matches 474; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSAAATGAPAPNQQPEVFCNQIFINNEHDAVSRRKTPPTVNPSTGEVICQVAGDKEDV 60
 DB :|||||
 QY 19 LSAATGAPAPNQQPEVFCNQIFINNEHDAVSRRKTPPTVNPSTGEVICQVAGDKEDV 76
 DB :|||||

A:Residues: 92-520 <EAF>
 A:Experimental source: liver
 R:Lee, J.E.; Cho, Y.D.
 Biochem. Biophys. Res. Commun. 189, 450-454, 1992
 A:Title: Purification and characterization of bovine brain gamma-aminobutyraldehyde dehydrogenase
 A:Reference number: PQ0543; MUID:53080596; PMID:1445496
 A:Accession: PQ0543
 A:Molecule type: protein
 A:Residues: 22-34 <LEE>
 A:Experimental source: brain
 C:Comment: This enzyme may be responsible for gamma aminobutyraldehyde dehydrogenase activity
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 F:1-21/Domain: transit peptide (mitochondrion) #status predicted <SIG>
 F:22-520/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAI>
 F:78-342/Domain: aldehyde dehydrogenase homology <ALDD>
 F:288-322/Active site: Glu, Cys #status predicted
 F:475/Binding site: NAD (Cys) #status predicted

Query Match 92.8% Score 2436; DB 1; Length 520;
 Best Local Similarity 91.8% Pred. No 2 90-180;
 Matches 457; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

QY 3 AAATQAVAPNQQPEVFCNOIFINNEWHDAVSRRKTPPTVNPSTGEVICOVAGKEDVDK 62
 DB 22 SAATQAVPTNQPEVLYNIOFINNEWHDAVSRRKTPPTVNPSTGEVICOVAGKEDVDK 81
 QY 63 AREGRCAGFOLGSPWRMDASHGRLLRLADLIERDRIYLALETLDNGKPYVLSYLD 122
 DB 82 AVKAAAFOLGSPWRMDASERGLRLADLIERDRIYLALETLDNGKPYVLSYLD 141
 QY 123 LDWVLKCLRYAGWADKYHGKTIPIDGFPSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 182
 DB 142 LDWVLKCLRYAGWADKYHGKTIPIDGFPSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 201
 QY 183 ALATGNVVMKVAEQPTLTALYVANLIKEAGPPGVNIVPGPGTAGAATASHEDVDK 242
 DB 202 ALATGNVVMKVAEQPTLTALYVANLIKEAGPPGVNIVPGPGTAGAATASHEDVDK 261
 QY 243 AFTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUAWAVEQAHAFFALFENQGC 302
 DB 262 AFTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUAWAVEQAHAFFALFENQGC 321
 QY 303 CCAGSRTEVEDIYDEEVSVARAKSRVVGNNPFDSTKTEGQGVDETOFKKILGINTCK 362
 DB 322 CCAGSRTEVEDIYAEFVSAPAKSRVVGNNPFDSTKTEGQGVDETOFKKILGINTCK 381
 QY 363 QERAKLIGGRTIADRGYFIQPTVFGVDQGMETAKEEIPGVVMQILKFKTIEEVVGRAN 422
 DB 382 FPGIKLGGGAAADPGYFIQPTVFGDLQGMETAKEEIPGVVMQILKFKSMEEVVGRAN 441
 QY 423 NSTYGLAAAVFTKDLKANYLSQALQAGTVVWVNCYDVFGAQQSPGGMSSSGRELGEY 482
 DB 442 NSTYGLAAAVFTKDLKANYLSQALQAGTVVWVNCYDVFGAQQSPGGMSSSGRELGEY 501
 QY 483 LGAYTEVKTVTVKVPQKN 500
 DB 502 LGAYTEVKTVTVKVPQKN 519

RESULT 5
 S00364
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2, mitochondrial horse (tentative sequence)
 C:Species: Equus caballus (domestic horse)
 C:Date: 30-Jan-1989 #sequence_revision 31-Jan-1997 #text_change 04-Jun-2002
 R:Johansson, J.; von Bahr-Lindstrom, H.; Jeck, R.; Woenckhaus, C.; Joernvall, H.
 Eur. J. Biochem. 172, 527-533, 1988

A:Title: Mitochondrial aldehyde dehydrogenase from horse liver: Characterizations of the S
 A:Reference number: S00364; MUID:88166740; PMID:3450012
 A:Accession: S00364
 A:Molecule type: protein
 A:Residues: 1-500 <JOH>
 A:Note: amino terminal residue is uncertain, 1 Ser and 1 Leu were also found; forms h
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta
 F:58-322/Domain: aldehyde dehydrogenase homology <ALDD>
 F:288-402/Active site: Glu, Cys #status predicted
 F:455/Binding site: NAD (Cys) #status predicted

Query Match 92.7% Score 2414; DB 1; Length 500;
 Best Local Similarity 92.6% Pred. No 3 90-180;
 Matches 452; Conservative 22; Mismatches 20; Indels 0; Gaps 0;

QY 2 SAATQAVAPNQQPEVFCNOIFINNEWHDAVSRRKTPPTVNPSTGEVICOVAGKEDVDK 61
 DB 1 AAATQAVAPNQQPEVFCNOIFINNEWHDAVSRRKTPPTVNPSTGEVICOVAGKEDVDK 60
 QY 62 KAPGPPGAFOLGSPWRMDASHGRLLRLADLIERDRIYLALETLDNGKPYVLSYLD 121
 DB 61 KAVKAAAFOLGSPWRMDASERGLRLADLIERDRIYLALETLDNGKPYVLSYLD 120
 QY 122 LDWVLKCLRYAGWADKYHGKTIPIDGFPSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 181
 DB 121 LDWVLKCLRYAGWADKYHGKTIPIDGFPSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 180
 QY 182 PALATGNVVMKVAEQPTLTALYVANLIKEAGPPGVNIVPGPGTAGAATASHEDVDK 241
 DB 181 PALATGNVVMKVAEQPTLTALYVANLIKEAGPPGVNIVPGPGTAGAATASHEDVDK 240
 QY 242 VAPTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUAWAVEQAHAFFALFENQGC 301
 DB 241 VAPTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUAWAVEQAHAFFALFENQGC 300
 QY 302 QYTAGSRTEVEDIYAEFVSAPAKSRVVGNNPFDSTKTEGQGVDETOFKKILGINTCK 361
 DB 301 CCAGSRTEVEDIYAEFVSAPAKSRVVGNNPFDSTKTEGQGVDETOFKKILGINTCK 360
 QY 362 KQERAKLIGGRTIADRGYFIQPTVFGVDQGMETAKEEIPGVVMQILKFKTIEEVVGRKA 421
 DB 361 KQERAKLIGGRTIADRGYFIQPTVFGVDQGMETAKEEIPGVVMQILKFKTIEEVVGRKA 420
 QY 422 NSTYGLAAAVFTKDLKANYLSQALQAGTVVWVNCYDVFGAQQSPGGMSSSGRELGEY 481
 DB 421 NSTYGLAAAVFTKDLKANYLSQALQAGTVVWVNCYDVFGAQQSPGGMSSSGRELGEY 480
 QY 482 LGAYTEVKTVTVKVPQKN 500
 DB 481 LGAYTEVKTVTVKVPQKN 499

RESULT 6
 A08872
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 5 precursor, mitochondrial human
 C:Species: Homo sapiens (man)
 C:Date: 27-Mar-1992 #sequence_revision 4-Jan-1997 #text_change 03-Jun-2002
 R:Hsu, L.C.; Chang, W.C.
 J. Biol. Chem. 266, 12267-12269, 1991
 A:Title: Cloning and characterization of a new tunnelled human aldehyde dehydrogenase
 A:Reference number: A08872; MUID:91266241; PMID:2061411
 A:Accession: A08872
 A:Molecule type: DNA
 A:Residues: 1-517 <HSU>
 A:Cross-references: GB:M64967; NID:q447184; PID:AAA6840.1; PID:q1264008
 A:Note: 86-Ala and 107-Leu were also found

C.Comment: Based on similarity, this form is assumed to be mitochondrial and to have pro

C:Gene: GDB:ALDH5; ALDHX

A:Cross-references: GDB:128788

A:Map position: 9p13-9p13

A>Note: the only info occurs before the initiator codon

C:Complex: homotetramer (probably)

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A>Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase

F:1-20/Domain: trans peptide (mitochondrial) #status predicted <NFI>

F:21-517/Product: aldehyde dehydrogenase (NAD⁺) #status predicted <NAT>

F:75-339/Domain: aldehyde dehydrogenase homology <ALDH>

F:285-319/Active site: Glu, Cys #status predicted

F:285-319/Active site: Glu, Cys #status predicted

Query Match 73.9%, Score 1939; DB 1; Length 517;

Best Local Similarity 72.9%; Pred. No. 7.2e-142;

Matches 364; Conservative 61; Mismatches 72; Indels 2; Gaps 1;

QY 2 SAATQAVPAPNQOPEVFCNOIFINNEHDAVSRTPTTVPSTGEVICOVAGKEDVD 61

DB 20 SAA--ALPSPILNPDPYNGFLINNEWDQAVSKTPTVNTTTEGIVHVAEGRADVD 77

QY 62 KAREGPPAPQJGSPWRPMDASHSGRLNRLADLTERDPTVLALETLDNGKPYVSYLV 121

DB 78 PAVKAARVPFLGSPWRPMDASERGLNRLADLVERDPTVLALETLDNGKPYVSYLV 137

QY 122 LDMVLKCLRYAGWADKYHCKTIPIDGDFSYTRHPEVGVCGGIIHWNPTLLMOAWKLG 181

DB 138 DLDEIVKRYVFAGWADKWHCKTIPMDGQHFCEITHPEPVGVCGGIIHWNPTLLMOAWKLG 197

QY 182 PALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNIVPGPPTAGAAIASHEDVDK 241

DB 198 PALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNIVPGPPTAGAAIASHEDVDK 257

QY 242 VAFSTGSLGRVLAAGAA--SSNLKRVTLLEAGKSPNIMSLALMIMWAVEAHAFEPNAG 301

DB 258 VAFSTGSLGVHLQKAA--TSNLKRVTLLEAGKSPNIVLAIAWAEHAFEPNAG 317

QY 302 CCAASRKLTVQDITVDEIVVRSVAFKSVGVGNFHSKTPGSPVAVLEPKKIDSYINT 361

DB 318 CCAASRKLTVQDITVDEIVVRSVAFKSVGVGNFHSKTPGSPVAVLEPKKIDSYINT 377

QY 362 KQESAKLLCGGERTGERGFFIKPIVFGSVQIIMPIAKKEIPGVPQPIKPKKTERVVEFA 421

DB 378 KQESAKLLCGGERTGERGFFIKPIVFGSVQIIMPIAKKEIPGVPQPIKPKKTERVVEFA 447

QY 422 NNSTYGLAAAVFTKDLKANYLSUALQAGTVVWVNCYDVPQAQSPFGGKNGSGRELSEY 481

DB 438 NNTYGLAAAVFTKDLKANYLSUALQAGTVVWVNCYDVPQAQSPFGGKNGSGRELSEY 497

QY 482 GLQAYTEVKTVTVKVPQKN 500

DB 498 GLKATIEVKITIKVPQKN 516

RESULT 7

S14624

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 1, cytosolic - chicken

C:Species: Gallus gallus (chicken)

C>Date: 21-Nov-1993 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002

C:Accession: S14624

C:Keywords: R;

A:Description: High levels of aldehyde dehydrogenase transcripts in the undifferentiated

A:Reference number: S14629

A:Accession: S14629

A:Molecule type: mRNA

A:Residues: 1-509 <ZNA>

A:Cross-references: EMBL X58869, NID 93012, F11N CAA41679.1, P1D 963013

A:Experimental source: retina

C:Complex: homotetramer

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A>Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase

F:57-331/Domain: aldehyde dehydrogenase homology <ALDH>

F:277-311/Active site: Glu, Cys #status predicted

F:464/Binding site: NAD (Cys) #status predicted

Query Match 68.6%, Score 1802; DB 1; Length 509;

Best Local Similarity 67.3%; Pred. No. 2.7e-131;

Matches 305; Conservative 96; Mismatches 96; Indels 0; Gaps 0;

QY 3 AAATQAVPAPNQOPEVFCNOIFINNEHDAVSRTPTTVPSTGEVICOVAGKEDVD 62

DB 11 APVLPALPEPLKDLKIKYIKIFINNEHDSVSKKEEVPNANLEKICEVAEGSKADIDK 70

QY 63 APGPPGAFOLGSPWRPMDASHSGRLNRLADLTERDPTVLALETLDNGKPYVSYLV 122

DB 71 AVKAAPKAFELGSPWRPMDASERGLNRLADLVERDPTVLALETLDNGKPYVSYLV 130

QY 123 LDMVLKCLRYAGWADKYHCKTIPIDGDFSYTRHPEVGVCGGIIHWNPTLLMOAWKLG 182

DB 141 LDMVLKCLRYAGWADKYHCKTIPIDGDFSYTRHPEVGVCGGIIHWNPTLLMOAWKLG 190

QY 183 ALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNIVPGPPTAGAAIASHEDVDK 242

DB 191 ALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNIVPGPPTAGAAIASHEDVDK 250

QY 244 AFTSTGSTRVTVAAAG--SSNLKRVTLLEAGKSPNIMSLALMIMWAVEAHAFEPNAG 302

DB 251 SETSGTEVGLIKKEAGKTNLKVTLLEAGKSPNITPADADLDEAFEAHGLIYHGQGC 310

QY 404 CCAASRKLTVQDITVDEIVVRSVAFKSVGVGNFHSKTPGSPVAVLEPKKIDSYINT 462

DB 411 CCAASRKLTVQDITVDEIVVRSVAFKSVGVGNFHSKTPGSPVAVLEPKKIDSYINT 470

QY 464 KQESAKLLCGGERTGERGFFIKPIVFGSVQIIMPIAKKEIPGVPQPIKPKKTERVVEFA 523

DB 471 KQESAKLLCGGERTGERGFFIKPIVFGSVQIIMPIAKKEIPGVPQPIKPKKTERVVEFA 530

QY 423 NNSTYGLAAAVFTKDLKANYLSUALQAGTVVWVNCYDVPQAQSPFGGKNGSGRELSEY 482

DB 431 NNTYGLAAAVFTKDLKANYLSUALQAGTVVWVNCYDVPQAQSPFGGKNGSGRELSEY 490

QY 483 GLQAYTEVKTVTVKVPQKN 500

DB 491 LQETIEVKITIKVPQKN 508

RESULT 8

S14624

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 - mouse

N:Alternate names: retinaldehyde-specific dehydrogenase

C:Species: Mus musculus (house mouse)

C>Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 03-Jun-2002

C:Accession: S74224

R:Zhao, D.; McCafferty, P.; Twiss, K.J.; Neve, R.L.; Hodan, P.; Chiu, W.W.; Braeger, U.

Fur. J. Biochem. 240, 15-22, 1996

A:Title: Molecular identification of a major retinoic-acid-synthesizing enzyme, a ret

A:Reference number: S74224; M01D063985; F01D063985

C:Accession: S74224

A:Molecule type: mRNA

A:Residues: 1-499 <ZNA>

A:Cross-references: EMBL X59274, NID 9314 <ZNA>, F11N CAA6766.1; P1D:gl430869

A:Experimental source: strain C3H/He; cell type embryonal carcinoma; cell line p19 te

C:Gene: RALDH-2

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: NAD; oxidoreductase

Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKKEAGFPFGVGNVIVPGYGPAGAAISSHHMD 240
 QY 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 Db 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 QY 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 Db 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 QY 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 Db 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVTKVPKN 500
 Db 481 YGFHEYTEVKTIVTKISOKN 500

RESULT 10

S14752

aldehyde dehydrogenase (NAD) (EC 1.2.1.3), cytosolic - sheep

C:Species: Ovis orientalis aries, ovis ammon aries (domestic sheep)

C:Date: 10-Mar-1997 #sequence_revision 13-Sep-1998 #text_change 03-Jun-2002

C:Accession: S78582; S14752

R: Tweedie, J.W.

submitted to the EMBL Data Library, July 1994

A:Reference number: S78582

A:Accession: S78582

A:Molecule type: mRNA

A:Residues: 1-501 <KIT>

A:Cross references: EMBL G12761, NID:J527681, PDB:AAA85435.1, ID:J527682

A:Experimental source: liver

R:Kitson, T.M.; Hill, J.P.; Midwinter, G.G.

Biochem J 275, 207-210, 1991

A:Title: Identification of a catalytically essential nucleophilic residue in sheep liver

A:Reference number: S14752, M010:91207293, PDB:2G18476

A:Accession: S14752

A:Molecule type: protein

A:Residues: 298-308 <KIT>

A:Experimental source: liver

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:59-323/Domain: aldehyde dehydrogenase homology <ALD>

F:302/Active site: Cys #status experimental

Query Match

Best Local Similarity 67.2%; Score 1763; DB 2; Length 501,

Matches 333; Conservative 67; Mismatches 100; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNNOQPEVFCNQIFINNEHDAVSRKFTPTVNPSTGEVICQVAGGKEDV 60
 Db 1 MSSAMPDVPAPLTNLQFYTKIFINNEHSSVSGKFPVPNPATEKLECEVEEGKADV 60
 QY 61 DKAREGRPAFQGLGSPWRMDASHGRLNLNADLIERDRTYLAALETLDNGKPYVISYL 120
 Db 61 DKAVKAAPQAFQGLGSPWMTASEGFLNKLADLIERDRTYLAALETLDNGKPYVISYL 120
 QY 121 VLDLWVLCRYAGWADKYHGKTIPIIDGDFEFSTRHEPVGVCQGIIPWNEPFLMGAWKL 180
 Db 121 MLDGGTKTKRYAGWADKYHGKTIPIIDGDFEFSTRHEPVGVCQGIIPWNEPFLMGAWKL 180
 QY 181 GPALATGNVNVKVAEQTPLTALXVANLIKKEAGFPFGVGNVIVPGYGPAGAAISSHHMD 240
 Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKKEAGFPFGVGNVIVPGYGPAGAAISSHHMD 240
 QY 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 Db 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 QY 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 Db 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 QY 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 Db 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVTKVPKN 500
 Db 481 YGFHEYTEVKTIVTKISOKN 500

RESULT 11

JC5553

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 03-Jun-2002

C:Accession: JC5553

R:Kathmann, E.C.; Lipsky, J.J.

Biochem Biophys Res Commun 236, 527-531, 1997

A:Title: Cloning of a cDNA encoding a constitutively expressed rat liver cytosolic al

A:Reference number: J5553; M010:91207293, PDB:1Q240474

A:Accession: JC5553

A:Molecule type: mRNA

A:Residues: 1-501 <KIT>

A:Cross references: PDB:1Q240474, NID:J527681, PDB:AAA85435.1, ID:J527682

A:Experimental source: liver

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: oxidoreductase

F:59-323/Domain: aldehyde dehydrogenase homology <ALD>

Query Match

Best Local Similarity 66.9%; Score 1757; DB 2; Length 501,

Matches 331; Conservative 69; Mismatches 100; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNNOQPEVFCNQIFINNEHDAVSRKFTPTVNPSTGEVICQVAGGKEDV 60
 Db 1 MSSAPQAPVAPLAPLANLKTQHTKIFINNEHDSVSGKFPVNPATEEIVCHVEEGKADV 60
 QY 61 DKAREGRPAFQGLGSPWRMDASHGRLNLNADLIERDRTYLAALETLDNGKPYVISYL 120
 Db 61 DKAVKAAPQAFQGLGSPWMTASEGFLNKLADLIERDRTYLAALETLDNGKPYVISYL 120
 QY 121 VLDLWVLCRYAGWADKYHGKTIPIIDGDFEFSTRHEPVGVCQGIIPWNEPFLMGAWKL 180
 Db 121 SPAGSIFKALRYAGWADKYHGKTIPIIDGDFEFSTRHEPVGVCQGIIPWNEPFLMGAWKL 180
 QY 181 GPALATGNVNVKVAEQTPLTALXVANLIKKEAGFPFGVGNVIVPGYGPAGAAISSHHMD 240
 Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKKEAGFPFGVGNVIVPGYGPAGAAISSHHMD 240
 QY 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 Db 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVAGQAHFALFNQ 400
 QY 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 Db 301 QCCAGSRTFVDEYDFVVRKSVARAKSKRVNRPFSKTRQSPQVDETFKKILGYINT 360
 QY 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 Db 361 GKQEGAKLLCGGIAADGNYFYIQTPTVFWGHTIAKEELFPHVMVILKFKETIEEVGP 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALAGTIVWVNCYDVFCAQSPFGYKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVTKVPKN 500
 Db 481 YGFHEYTEVKTIVTKISOKN 500


```
Db 361 GRKGAKECGGGWGNKGFVQTVFSNVTDEMTAKELFGVQVQIMKFKSLDDVIKR 420
QY 421 ANNSTYGLAAAVFTKDDKANYLSQALQAGTVVWVNCYDVFGAQSPPFGYKMSGSGRELGE 480
Db 421 ANNSTYGLAAVFTKDDRAITVSSALQAGVWVWVNCYMLISAQCPFGGFKMSGNGRELGE 480
QY 481 YGLQAYTEVKTVTKVPKKN 500
Db 481 HGLYETELKIVAMKISQKN 500

RESULT 14
S02302
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 1, cytosolic - horse
C:Species: Equus caballus (domestic horse)
C:Date: 01-Dec-1989 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: S02302
R:Von Bahr-Lindstrom, H.; Hempel, J.; Joernvall, H.
A:Title: The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase. Relationship t
A:Reference number: S02302; MUID:84208025; PMID:6723662
A:Accession: S02302
A:Molecule type: protein
A:Residues: 1-500 :VAN-
A:Note: residues in the regions 1-16 and 443-475 were positioned by homology with the hu
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase; saliva
F:780-334/Domain: aldehyde dehydrogenase homology <ALDH>
F:1/Modified site: blocked amino end (Ser) (probably acetylated) #status experimental
F:268,302/Active site: Glu, Cys #status predicted
F:455/Binding site: NAD (Cys) #status predicted

Query Match 66.1%; Score 1736; DB 1; Length 500;
Best Local Similarity 66.1%; Pred. No. 3.3e-126;
Matches 330; Conservative 65; Mismatches 104; Indels 0; Gaps 0;

QY 2 SAATQAPAPNQOPEVFCNCFINNEWHDAVSRKTFPTVNPSTGEVICOVAGKEDVD 61
Db 1 SSGTPTDPLVLTIDTKFYTKIFINNEWHDSVSGKFPVNPNAETKLECEVEGKEDVN 60
QY 62 KAREGPAGFALGSPWRMDASHSGRLNRLADLIERDRTYLALETDLONGKPYVTSYLV 121
Db 61 KAVAAARCAFGISGSPWPTMDASERGLLYKLADIVERDRLILATMESMNGGLFSNAYLM 120
QY 122 DLDWVYKLRYYAGWADKYHGKTTIPDGDFSYTRHEPVGCGQIIPWNPPLLMQAWKLG 181
Db 121 DLGGGLKTLRYCAGWARKIQGPTIPSDGNFTYTRHEPVGCGQIIPWNPPLLMFLWKTA 180
QY 182 PALATGNVVMKVAEQTPLTALYVANLIKEAGFPFGVNVIVPGFGPTAGAAATASHEDVDK 241
Db 181 PALSQGNVVMKVAEQTPLTALYVANLIKEAGFPFGVNVIVPGFGPTAGAAATASHEDVDK 240
QY 242 VAFTGSTIEGRVIOVAAGSSNLKRVTLLEGGKSPNIIIMSDADMVAEQAHAFLFPNQOQ 301
Db 241 VAFTGSTIEGRVIOVAAGSSNLKRVTLLEGGKSPNIIIMSDADMVAEQAHAFLFPNQOQ 300
QY 302 CCCAGSRFTVQEDYIDFVVRVARAKSRVVGVPFUSKTEGQVDETQFKKILGYNTG 361
Db 301 CCCAASRFTVQEDYIDFVVRVARAKSRVVGVPFUSKTEGQVDETQFKKILGYNTG 360
QY 362 KOEGAKLLGGGIAADRGYFTQPTVFGVDGDMGTIAKEEIFGPMVQILKFKTIEEVRANNSTYG 421
Db 361 KKEGAKLECGGPGWNGKGYFTQPTVFSNVSDMKIAKEELFGVQVQIMKFKSLDDVIKR 420
QY 422 NNSTYGLAAAVFTKDDKANYLSQALQAGTVVWVNCYDVFGAQSPPFGYKMSGSGRELGEY 481
Db 421 NNSTYGLFAGSFTKDDKAITVSSAALQAGTVVWVNCYGVVSAQCPFGGFKMSGNGRELGEY 480
```

```
QY 482 GLOAYTEVKTVTKVPKKN 500
Db 481 GFHEYTEVKTVTKISQKN 499

RESULT 15
A55684
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 6 precursor, salivary - human
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1995 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: A55684
R:Hsu, L C; Chang, W C; Hiraoka, L; Hsieh, C I.
Genomics 24, 333-341, 1994
A:Title: Molecular cloning, genomic organization, and chromosomal localization of an
A:Reference number: A55684; MUID:95213025; PMID:7698756
A:Accession: A55684
A:Molecule type: mRNA; DNA
A:Residues: 1-512 <HSU>
A:Cross-references: GB:U07919; NID:g995897; PIDN:AAA79036.1; PID:g544482
A:Comment: This isozyme is found at highest levels in saliva, stomach, and kidney and
C:Genetics:
A:Gene: GDB:ALDH6
A:Cross-references: GDB:364103; OMIM:600463
A:Map position: 15q26-15q26
A:Initiator: 10/1; 18/1; 15/1; 15/1; 17/1; 22/1; 24/1; 25/1; 41/1; 44/1;
C:Function:
A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase; saliva
F:780-334/Domain: aldehyde dehydrogenase homology <ALDH>
F:280,314/Active site: Glu, Cys #status predicted
F:467/Binding site: NAD (Cys) #status predicted

Query Match 65.4%; Score 1718; DB 1; Length 512;
Best Local Similarity 65.5%; Pred. No. 8.4e-125;
Matches 323; Conservative 66; Mismatches 104; Indels 0; Gaps 0;

QY 8 AVAPNQOPEVFCNCFINNEWHDAVSRKTFPTVNPSTGEVICOVAGKEDVDKAREGR 67
Db 19 ALPPIRLNLEVKFTKIFINNEWHESKSGKFAFCNSTREICEVEGDKPDVQKAVEAA 78
QY 68 PGACQLASPPMPMPASHSGPLINPLADLIEPQPTYLALETDLNGKPYVTSYLVLDMDVL 127
Db 79 QVAFQSGSPWPLDALSPGLIHQIADIVRFDATIAALFTMDYGRPFTHAEPIDLEGGCI 138
QY 128 KCLRYAGWADKYHGKTTIPDGDFSYTRHEPVGCGQIIPWNPPLLMQAWKLGSPALATG 187
Db 139 RLLRYAGWADKYHGKTTIPDGDFSYTRHEPVGCGQIIPWNPPLLMQAWKLGSPALATG 198
QY 188 NVVMKVAEQTPLTALYVANLIKEAGFPFGVNVIVPGFGPTAGAAATASHEDVDKVAFTGS 247
Db 199 NTMVLKPAEQPLTALYVLSLIKEAGFPFGVNVIVPGFGPTAGAAATASHEDVDKVAFTGS 258
QY 248 TEIGRVIOVAAGSSNLKRVTLLEGGKSPNIIIMSDADMVAEQAHAFLFPNQOQCCAGS 307
Db 259 TEVGLKVEASNLKRVTLLEGGKSPNIIIMSDADMVAEQAHAFLFPNQOQCCAGS 318
QY 308 KTFVQURPFDYFVVRVARAKSRVVGVPFUSKTEGQVDETQFKKILGYNTGKUEGAK 367
Db 319 RVFVEQVYSEFVVRVARAKSRVVGVPFUSKTEGQVDETQFKKILGYNTGKUEGAK 378
QY 368 LLCGGIAADRGYFTQPTVFGVDGDMGTIAKEEIFGPMVQILKFKTIEEVRANNSTYG 427
Db 379 LECGGIAMEDKGLFKTFVSEVTDNMRIAKEEIFGPMVQILKFKTIEEVRANNSTYG 438
QY 428 LAAAVFTKDDKANYLSQALQAGTVVWVNCYDVFGAQSPPFGYKMSGSGRELGEYGLQAYT 487
Db 439 LTAAVFTKDDKANYLSQALQAGTVVWVNCYDVFGAQSPPFGYKMSGSGRELGEYGLQAYT 498
QY 488 EVKTVTKVPKKN 500
Db 499 EVKTVTKIGDKN 511
```

Search completed: June 24, 2003, 10:18:30
Job time : 20.1339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:55 ; Search time 10.0735 seconds
(without alignments)
2059.300 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAATGAVPAPNDQPEVF YGLQAYIEVKIVTKVPQKN 500

Scoring table: BLOSUM62

Gapop 10 0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match %	Length	ID	Description
1	2572	98.0	517	1	DHAM_HUMAN
2	2563	95.4	500	1	DHAM_MSAU
3	2500	95.2	519	1	DHAM_RAT
4	2480	94.5	519	1	DHAM_MOUSE
5	2436	92.8	520	1	DHAM_BOVIN
6	2434	92.7	500	1	DHAM_HORSE
7	1929	73.5	517	1	DHAM_HUMAN
8	1802	68.6	509	1	DHAM_CHICK
9	1795	68.4	499	1	DHAM_RAT
10	1794	68.3	499	1	DHAM_MOUSE
11	1790	68.2	499	1	DHAM_HUMAN
12	1771	67.5	499	1	DHAM_CHICK
13	1769	67.4	500	1	DHAM_HUMAN
14	1758	67.0	500	1	DHAM_SHEEP
15	1752	66.7	500	1	DHAM_MOUSE
16	1752	66.7	500	1	DHAM_RAT
17	1751	66.7	500	1	DHAM_BOVIN
18	1736	66.1	500	1	DHAM_HORSE
19	1718	65.4	512	1	DHAM_HUMAN
20	1711	65.2	500	1	DHAM_RAT
21	1706	65.0	501	1	DHAM_MOUSE
22	1697	64.6	501	1	DHAM_ELEPH
23	1587	60.5	497	1	DHAM_ENCAP
24	1526	58.2	494	1	DHAM_MOUSE
25	1499	57.1	495	1	DHAM_MOUSE
26	1425.5	54.3	497	1	DHAM_HUMAN
27	1422.5	54.2	511	1	DHAM_YEAST
28	1353.5	51.6	497	1	DHAM_ASPIG
29	1305.5	49.7	496	1	DHAM_CHICK
30	1264	48.3	495	1	DHAM_HUMAN
31	1258	47.9	500	1	DHAM_HUMAN
32	1235.5	47.1	519	1	DHAM_YEAST
33	1199.5	45.7	498	1	DHAM_HUMAN

RESULT 1	ID	DHAM_HUMAN	STANDARD	PRT	517 AA
AC	P05091	003639			
DT	13-AUG-1987	(Rel. 05, Created)			
DT	01-JAN-1990	(Rel. 13, Last sequence update)			
ET	15-JUN-2002	(Rel. 41, Last annotation update)			
EF	Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH)				
DE	class 2) (ALDH1) (ALDH-E2).				
GN	ALDH2 OR ALDH.				
OS	Homo sapiens (Human).				
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;				
OC	Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88256152; PubMed=3838413;				
RA	Hsu L.C., Bendel R.E., Yoshida A.;				
RT	"Genomic structure of the human mitochondrial aldehyde dehydrogenase				
RT	gene.";				
RL	Genomics 2:57-65(1988).				
RN	[2]				
RP	SEQUENCE OF 1A-517				
RC	TISSUE=Liver;				
RX	MEDLINE=86055846; PubMed=4065146;				
RA	Hempel J., Kaiser R., Joernvall H.;				
RT	"Mitochondrial aldehyde dehydrogenase from human liver. Primary				
RT	structure, differences in relation to the cytosolic enzyme, and				
RT	functional correlations.";				
PL	Eur. J. Biochem. 153:13-28(1985).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Muscle;				
RX	MEDLINE=87174836; PubMed=3562250;				
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;				
RT	"Isolation and sequence analysis of a full length cDNA clone coding				
RT	for human mitochondrial aldehyde dehydrogenase.";				
PL	Nucleic Acids Res. 15:3170-3179(1987).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Muscle;				
RX	MEDLINE=87219091; PubMed=3583651;				
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;				
RT	"Evidence for a signal peptide at the amino terminal end of human				
RT	mitochondrial aldehyde dehydrogenase.";				
PL	FEBS Lett. 215:233-236(1987).				
RN	[5]				
RP	PEPTIDES TO N-TERMINUS.				
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;				
RT	"Evidence for a signal peptide at the amino terminal end of human				
RT	mitochondrial aldehyde dehydrogenase.";				
PL	FEBS Lett. 215:233-236(1987).				
RN	[6]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lymph;				
RA	Straussberg R.;				
RT	Submitted (FEF-2001) to the EMBL/GenBank/DBJ databases.				
RN	[7]				

P40047 saccharomyc
P75891 homo sapien
P28937 rattus norv
P54115 saccharomyc
P47771 saccharomyc
P54114 saccharomyc
P71016 bacillus su
P28237 beta vulgar
P17202 spinacia ol
P42757 atiplex ho
P17445 escherichia
P23883 escherichia

Good an email to license@lsb-sib.ch).

or send an email to license@iso-sid.org.					
CC	EMBL; M20456;	AAA51693.1;	JOINED.		
CC	DR	EMBL; M20444;	AAA51693.1;	JOINED.	
CC	EMBL; M20445;	AAA51693.1;	JOINED.		
DR	EMBL; M20446;	AAA51693.1;	JOINED.		
DR	EMBL; M20447;	AAA51693.1;	JOINED.		
DR	EMBL; M20448;	AAA51693.1;	JOINED.		
DR	EMBL; M20449;	AAA51693.1;	JOINED.		
DR	EMBL; M20450;	AAA51693.1;	JOINED.		
DR	EMBL; M20451;	AAA51693.1;	JOINED.		
DR	EMBL; M20452;	AAA51693.1;	JOINED.		
DR	EMBL; M20453;	AAA51693.1;	JOINED.		
DR	EMBL; M20454;	AAA51693.1;	JOINED.		
DR	EMBL; X05409;	CXA28590.1;	-		
DR	EMBL; Y00109;	CAA68290.1;	-		
DR	EMBL; BC002967;	AAC62967.1;	-		
DR	EMBL; K03001;	AA559500.1;	-		
DR	EMBL; M26760;	AAA51694.1;	-		
DR	EMBL; M54931;	AAAG2825.1;	ALT_FRAME.		
DR	EMBL; S00804;	DEH0E2.			
DR	PIR; A29975;	A29975.			
DR	PIR; A29975;	A29975.			
DR	PDB; 1CW3;	10-JAN-00			
DR	Genew; HGNC:404;	ALDH2.			
DR	MIM; 100650;	-			
DR	InterPro; IPK002086;	Aldehyde_dehydr.			
DR	Pfam; PF00171;	aldehyd; 1.			
DR	PROSITE; PS00070;	ALDEHYDE_DEHYDR_CYS; 1.			
DR	PROSITE; PS00687;	ALDEHYDE_DEHYDR_GLD; 1.			
DR	Oxidoreductase, NAD;	Mitochondrion; Trans it peptido; Polymorphism;			
KW	3D-structure.	1			
KW	TRANSIT	17			
FT	CHAIN	18	517		
FT	NP_BIND	262	267		
FT	ACT_SITE	285	285		
FT	ACT_SITE	419	419		
FT	VARIANT	337	337		
FT	VARIANT	496	496		
FT	VARIANT	504	504		
FT	VARIANT	504	504		
FT	CONFLICT	7	12		
FT	CONFLICT	7	26		
FT	CONFLICT	18	18		
FT	CONFLICT	80	85		
FT	CONFLICT	216	216		
FT	CONFLICT	218	218		
FT	CONFLICT	247	247		
FT	CONFLICT	380	380		
FT	CONFLICT	462	463		
FT	SEQUENCE	517 AA;	56381 MW;		

Query Match	Score	Indels	Gaps
Best Local Similarity	Pred. No. 4, 7c-190		
Matches	2; Mismatches	7; Indels	0; Gaps
1	MSAAATQAVPAPNQQPEVPCNGIFINNEWHJHVAVSRRKTFPTVNFSTGEVITCVVAEGDKEDW	60	
17	LSAAATQAVPAPNQQPEVPCNGIFINNEWHJHVAVSRRKTFPTVNFSTGEVITCVVAEGDKEDW	76	
61	DKAREGRPGAFQLGSGPWRRGMDASHGRLINKLAULDERDFTYLAALFETLNLKPFVVSYL	120	
77	DKAKAAHAAAFQLGSGPWRRGMDASHGRLINKLAULDERDFTYLAALFETLNLKPFVVSYL	146	
121	VLDLWVILKCLRYAGWADKYHGKTIPIIDGDFESYTRHEPVCVCQTLLPNWFETLMLAKKL	180	
137	VLDLWVILKCLRYAGWADKYHGKTIPIIDGDFESYTRHEPVCVCQTLLPNWFETLMLAKKL	196	
181	GPALATGVVWVWVAEQITPLTATAYVANLLTFACGPPGVVNIIVPGCGGTACAAATASHEFDV	240	

SEQUENCE OF 119-517 FROM N A
TISSUE=Liver;
MEDLINE=85216574; PubMed-2987944;
Hsu L.-C., Tani K., Fujiyoshi T., Kurachi K., Yoshida A.;
"Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2";
Proc. Natl. Acad. Sci. U.S.A. 82:3771-3775(1985).
[8]
SEQUENCE OF 119-517 FROM N A
MEDLINE=85252089; PubMed-4015823;
Yoshida A., Ikawa M., Hsu L.-C., Tani K.;
"Molecular abnormality and cDNA cloning of human aldehyde
dehydrogenases";
Alcohol 2:103-106(1985).
[9]
SEQUENCE OF 214-500 FROM N A.
TISSUE=Liver;
MEDLINE=87279033; PubMed-3610592;
Agarwal D.P., Goedde H.W.;
"Human aldehyde dehydrogenase isozymes and alcohol sensitivity.";
Isozymes Curr. Top. Biol. Med. Res. 16:21-48(1987).
[10]
DESCRIPTION OF ORIGIN OF CONFLICTS BETWEEN REF.2 AND DNA SEQUENCES.
MEDLINE=88005159; PubMed-3653404;
Hempel J., Hoeseg J.-O., Joernvall H.;
"Mitochondrial aldehyde dehydrogenase. Homology of putative targeting
sequence to that of carbamyl phosphate synthetase I revealed by
correlation of cDNA and protein data.";
FEBS Lett. 222:95-98(1987).
[11]
VARIANT LYS-504.
MEDLINE=84119449; PubMed-6582480;
Yoshida A., Huang I.-Y., Ikawa M.;
"Molecular abnormality of an inactive aldehyde dehydrogenase variant
commonly found in Orientals.";
Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).
[12]
VARIANT LYS-496.
MEDLINE=96119362; PubMed-8561277;
Novorodovsky A., Tsai S.-J., Goldfarb L., Peterson P., Long J.C.,
Goldman D.;
"Mitochondrial aldehyde dehydrogenase polymorphism in Asian and
American Indian populations: detection of new ALDH2 alleles";
Alcohol. Clin. Exp. Res. 19:1105-1110(1995).
[13]
X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).
MEDLINE=20095857; PubMed-10631996;
Ni L., Zhou J., Hurley T.D., Weiner H.;
"Human liver mitochondrial aldehyde dehydrogenase: three-dimensional
structure and the restoration of solubility and activity of chimeric
forms";
Protein Sci. 8:2784-2790(1999).
-1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
-1- PATHWAY: Ethanol utilization; second step.
-1- SUBUNIT: HOMOTETRAMER.
-1- SUBCELLULAR LOCATION: Mitochondrial matrix.
-1- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN
ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS,
IS BEING DIRECTLY ATTRIBUTED TO AN ENZYMATICALLY IMPAIRED ALDH
ISOZYME. THE INACTIVE VARIANT (ALLELE 2 OR ALDH2*2) IS DUE TO
SINGLE AMINO ACID EXCHANGE.
-1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
-1- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF.1
AND REF.2 AND THAT IN REF.3 AND 4 ARE PROBABLY ALL DUE TO
FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF.5 AND REF.6;
THE SEQUENCE DESCRIBED IN REF.9 DIFFERS FROM THAT SHOWN DUE TO TWO
FRAMESHIFTS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

Db 197 GPALATGNVVMKVAEOTPLTALYVANLIKEAGPPGVVNLVPGFGPAGAAIASHEVDV 256
QY 241 KVAFTGSTEIGRVTOVAGSSNLKRVTLLEGGKSPNIIIMSDADMWAVEQAHFALFFNQ 300
Db 257 KVAFTGSTEIGRVTOVAGSSNLKRVTLLEGGKSPNIIIMSDADMWAVEQAHFALFFNQ 316
QY 301 QCCAGSRTFVOEDIDYDFEVRSVARAKSRVGVNPFDSKTEOGQVQVDETOFKKILGYINT 360
Db 317 QCCAGSRTFVQETIYDFEVRSVARAKSRVGVNPFDSKTEOGQVQVDETOFKKILGYINT 376
QY 361 GKQEGAKLLCGGGIAADRGYFIQPTVFGDVJUGUMIAKEEIPFGVMOLKFKETIEEVVGR 420
Db 377 GKQEGAKLLCGGGIAADRGYFIQPTVFGDVJUGUMIAKEEIPFGVMOLKFKETIEEVVGR 436
QY 421 ANNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 480
Db 437 ANNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 496
QY 481 YGLQAYTEVKTIVTKVPQKN 500
Db 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 2
DHAM_MESAU
ID DHAM_MESAU STANDARD: PPT: 500 AA
AC P81178:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3) (ALDH class 2)
DE (ALDH1) (ALDH-E2).
GN ALDH2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE.
RC TISSUE=Liver;
PX MEDLINE=98034175; PubMed=9369242;
RA Hjeltnist L., Lundgren R., Norin A., Joernvall H., Vallee B.,
RA Kiyosaki A., Keung W.M.;
ET "Class 2 aldehyde dehydrogenase. Characterization of the hamster
ET enzyme, sensitive to dieldrin and conserved within the family of
ET multiple forms."
RL FEBS Lett. 416:99-102(1997)
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> H(2)O + an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.
CC -|- SUBUNIT: HOMOTETRAMER
CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR HSP: P05091; ICR3
DR InterPro: IPR002086; Aldehyde_dehyd.
DR Pfam: PF00171; aldh-dh; 1
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_G10; 1
DR PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1
KW Oxidoreductase; NAD; Mitochondrion.
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
SQ SEQUENCE 500 AA: 54334 MW: 8881866A04F493F CRC64:

Query Match 95.4%; Score 2503; DB 1; Length 500;
Best local similarity 95.0%; Pred. No. 1,6e-184;
Matches 474; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

QY 2 SAAATGAPAPNQGFEVFTNIFINNWHIAVSKHETPTVNPSTGEVIVQVABGKEDVD 61
Db 1 SAAATGAPAPNQGFEVFTNIFINNWHIAVSKHETPTVNPSTGEVIVQVABGKEDVD 60
QY 62 KAPETPPCAFQSGSPWPMIAWSHSTFLLNPLAFILFEPQYLAALFLLUNGKPYVISLV 121

```

```

Db 51 KAVFAARAAAFOLGSPWPMIAWSHSTFLLNPLAFILFEPQYLAALFLLUNGKPYVISLV 120
QY 122 DLDMY: KCLPYVAGWAKYBKETIPIGDFEFSTYTPHEFVCGQCIIPWNPPLIMQAWPLG 181
Db 121 DLDMY: KCLRYVAGWAKYBKETIPIGDFEFSTYTPHEFVCGQCIIPWNPPLIMQAWPLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKEAGPPGVVNLVPGFGPAGAAIASHEVDV 241
Db 181 PALATGNVVMKVAEOTPLTALYVANLIKEAGPPGVVNLVPGFGPAGAAIASHEVDV 240
QY 242 VAFPGSTEIGRVTOVAGSSNLKRVTLLEGGKSPNIIIMSDADMWAVEQAHFALFFNQ 301
Db 241 VAFPGSTEIGRVTOVAGSSNLKRVTLLEGGKSPNIIIMSDADMWAVEQAHFALFFNQ 300
QY 302 CCCAGSRTFVOEDIDYDFEVRSVARAKSRVGVNPFDSKTEOGQVQVDETOFKKILGYINT 361
Db 301 CCCAGSRTFVOEDIDYDFEVRSVARAKSRVGVNPFDSKTEOGQVQVDETOFKKILGYINT 360
QY 362 KQCAKTIICGGGIAADRGYFIQPTVFGDVJUGUMIAKEEIPFGVMOLKFKETIEEVVGR 421
Db 361 KQCAKTIICGGGIAADRGYFIQPTVFGDVJUGUMIAKEEIPFGVMOLKFKETIEEVVGR 420
QY 422 NNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 481
Db 421 NNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 480
QY 482 YGLQAYTEVKTIVTKVPQKN 500
Db 481 YGLQAYTEVKTIVTKVPQKN 499

RESULT 3
DHAM_RAT
ID DHAM_RAT STANDARD: PPT: 519 AA.
AC P11884:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15 JUN 2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
DE class 2) (ALDH1) (ALDH-E2).
GN ALDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=89210865; PubMed=2540003;
RA Farres J., Guan K.-L., Weiner H.;
RA Farres J., Guan K.-L., Weiner H.;
ET "Primary structures of rat and bovine liver mitochondrial aldehyde
ET dehydrogenases deduced from cDNA sequences."
RL Eur. J. Biochem. 180:57-74(1989).
RN [2]
RP SEQUENCE OF 1-29 FROM N.A.
RC TISSUE=Liver;
RC MEDLINE=89144717; PubMed=3442060;
RA Farres J., Guan K.-L., Weiner H.;
RA Farres J., Guan K.-L., Weiner H.;
ET "Sequence of the signal peptide for rat liver mitochondrial aldehyde
ET dehydrogenase."
RL Biochem. Biophys. Res Commun 150:1083-1087(1988).
RN [3]
RP SEQUENCE OF 1-19.
RC TISSUE=Liver;
RC MEDLINE=91378548; PubMed=1898068;
RA Jeng J., Weiner H.;
RA Jeng J., Weiner H.;
ET "Purification and characterization of catalytically active precursor
ET of rat liver mitochondrial aldehyde dehydrogenase expressed in
ET Escherichia coli."
FL Arch. Biochem. Biophys. 289:214-222(1991).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.

```

CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC

DR EMBL: X14977; CAA33101.1; -;
 DR EMBL: M19030; AAA40719.1; -;
 DR PIR: A27713; A27713;
 DR PIR: S03564; S03564;
 DR PIR: S17492; S17492;
 DR HSSP: P05091; ICW3;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1;
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1;
 KW Oxidoreductase; NAD; Acetylation; Mitochondrion; Transit peptide.
 FT TRANSIT 1 19 MITOCHONDRION.
 FT CHAIN 20 519 ALDEHYDE DEHYDROGENASE.
 FT MOD_RES 20 20 ACETYLATION (PROBABLE).
 FT NP_BIND 264 269 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 287 287 BY SIMILARITY.
 FT ACT_SITE 321 321 BY SIMILARITY.
 SQ SEQUENCE 519 AA: 56498 MW: 750748202 FL333E5 CRC64:

Query Match 95.2%; Score 2530; DB 1; Length 519;
 Best Local Similarity 94.8%; Prod. No. 2 9e-184;
 Matches 474; Conservative 12; Mismatches 14; Indels 0; Gaps 0.

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTEPTVNPSTGEVICOVAGDKEDV 60
 Db 19 LSAAATSAVPAPNQPEVFCNQIFINNEHDAVSRKTEPTVNPSTGEVICOVAGDKEDV 78
 QY 61 DKAFPGPPGAFQICSPRRMDASHGRLNRLANLEPTPTLALEPTIDNCKPYVSYL 120
 Db 79 DKAVKAAQAQAFQSGPWRMDASDRGLLYKLADLIERLRYLALEILDNCKPYVSYL 138
 QY 121 VLDLMLKCLRYAGWADYHGKTIPIDGDFEFTYRHEPVGCGQIIPWNEFLLMQAWKL 180
 Db 139 VLDLMLKCLRYAGWADYHGKTIPIDGDFEFTYRHEPVGCGQIIPWNEFLLMQAWKL 198
 QY 181 GPALATGNVVMKVAFTQPI TALYVANLIKEAGFPVGVNIVPGFPTAGAAIASHEDVD 240
 Db 199 GPALATGNVVMKVAFTQPI TALYVANLIKEAGFPVGVNIVPGFPTAGAAIASHEDVD 258
 QY 241 KVAFTSGTEIGRVIVAAAGSSNLKRVTLGLGCKSPNITMSUADMDWAVEQAHEALFFNQG 400
 Db 259 KVAFTSGTEVGHILQVAAAGSSNLKRVTLGLGCKSPNITMSUADMDWAVEQAHEALFFNQG 318
 QY 301 QCCACGSRTEVQEDIVDEFEVRSVARAKSRVGVNFGFUSKTEGCPVQVPTQPKILGYINT 460
 Db 319 QCCACGSRTEVQEDIVDEFEVRSVARAKSRVGVNFGFUSKTEGCPVQVPTQPKILGYILKS 378
 QY 361 GKVESAKLLTGGTAAFGGYFIQPTVFGSLVQDSMTIAKEEIPFGVMQLKFKTKTEEVVGR 420
 Db 379 GQCEGAKLLCGGGAADRGYFIQPTVFGDKDGMTIAKEEIPFGVMQLKFKTKTEEVVGR 438
 QY 421 ANNSTYGLAAAFIKULUKANYLSUALQAGIIVWNCYVFGAQSPPGKMGSGSGRELIGE 480
 Db 439 ANNKSYGLAAAFVTKDKANYLSUALQAGTVMINCYVFGAQSPPGKMGSGSGRELIGE 498
 QY 481 YGLQAYTEKVTIVTKVPQKN 500
 Db 499 YGLQAYTEKVTIVTKVPQKN 518

RESULT 4

UHAM_MOUSE
 ID UHAM_MOUSE STANDARD: PRT: 519 AA.
 AC P47738;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
 DE class 2) (AHD-M1) (ALDH1) (ALDH E2).
 GN ALDH2 OR AHD1 OR AHD-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE: 95047445; PubMed: 7958964;
 RA Chang C., Yoshida A.;
 FT "Cloning and characterization of the gene encoding mouse
 FT mitochondrial aldehyde dehydrogenase";
 RL Gene 148:331-336(1994).
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94335908; PubMed 8058082;
 RA Chen M., Achkar C., Gudas L.J.;
 FT "Enzymatic conversion of retinaldehyde to retinoic acid by cloned
 FT murine cytosolic and mitochondrial aldehyde dehydrogenases";
 RL Mol. Pharmacol. 46:88-96(1994).
 CC -1- FUNCTION: IS CAPABLE OF CONVERTING RETINALDEHYDE TO RETINOIC ACID.
 CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC -1- PATHWAY: Ethanol utilization; second step.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- INDUCTION: By retinoic acid; 4-5 fold increase.
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC

Query Match 94.5%; Score 2480; DB 1; Length 519;
 Best Local Similarity 94.2%; Prod. No. 1e-182;

Matches 471; Conservative 13; Mismatches 16; Indels 0; Gaps 0;

QY	1	MSAAATQAVPAPNOQPEVFCNOJPTINNEWHDAVSRKTEFTPTVNSIGEVICQVBAQKUEV	60
Db	19	LSAAATSAVPAPNIHQPEVFCNOJPTINNEWHDAVSRKTEFTPTVNSIGEVICQVBAQKNDK	78
QY	61	DKAPGPPGATQJASPPRRMCASHSGPLINPLACLTERPDTYLAALFTLNGKPPVIVSL	120
Db	79	DKAVKAAARAAALQGLSPWRMDASDRGLIYPLADLERPDTYLAALFTLNGKPPVIVSL	138
QY	121	VLDLMDVLKCLKAYACWALKYHKRTIPIGSGFFSYTFRHVGVSQTIPWNPFLIMCAWKL	188
Db	139	VLDLMDVLKCLRYAYAGWADKYHGKTIPIGSGFFSYTFRHVGVSQTIPWNPFLIMCAWKL	198
QY	181	GHALATGNVVMYKVAFOPTPLTATYVANLKEAGFPFPCVNNIPGFGPTAGAAIASHRDV	240
Db	199	GHALATGNVVMYKVAECTPLTATYVANLKEAFPPFCVNNIPGFGPTAGAAIASHREVP	258
QY	241	KVAFIIGSTELGPIVLCVAARSSNLKPVTLRIIGSGSPNIMSDMDWVAFQAHFALFFNQG	300
Db	259	KVAFITSTTEVGRILQVAAASSNLKPVTLRIIGSGSPNIMSDMLWVAFQAHFALFFNQS	318
QY	301	QVQVAGSPTFWCEDIYIEFVYKRSVAPAKSHVVNPNPUSKTEQSPQVDETQPKKILGYLNT	360
Db	319	QVQVAGSKTFWQENVYIEFVYKRSVAPAKSHVVNPNPUSPEIQSPQVDETQPKKILGYTKS	378
QY	361	KQGEKALIGSGSTAAALHRYFLPTVPSNVGQDMPTIAKBEIPGPMOILLKPTTIEVVR	420
Db	379	QOGEKALLCGGGAADRGVFIQPTVFGDKVKGDMIAKBEIPGPMOILLKPTTIEVVR	438
QY	421	NNNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNYDVEGASPSGSGYKMSGSISREJCE	480
Db	439	ANDSKYGLAAAVFTKDLKANYLSQALQAGTVMWNYDVEGASPPGGYKMSGSISREJCE	498
QY	481	YGLQAYTEVKTVTVKVPQKN	500
Db	499	YGLQAYTEVKTVTVKVPQKN	518

```

RESULT 5
DHAM_BOVIN
ID DHAM_BOVIN STANDARD PRI: 520 AA.
P20000:
DT DT 01-FEB-1991 (Rel 17, Created)
DT DT 01-FEB-1991 (Rel 17, Last sequence update)
DT DT 15-JUN-2002 (Rel 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
DE class 2) (ALDH1) (ALDH-E2).
DE GN ALDH2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Euminantia; Perissodactyla;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]
PP SEQUENCE FROM N.A.
RN TISSUE=Liver;
RX MEDLINE=90174194; PubMed=169944.
RA Guan K., Weiner H.;
RT "Sequence of the precursor of bovine liver mitochondrial aldehyde
RT dehydrogenase as determined from its cDNA, its gene, and its
RT functionality.";
RL Arch. Biochem. Biophys. 277:351-360(1990).
RN [2]
RP SEQUENCE OF 92-520 FROM N.A.
RX MEDLINE=89210865; PubMed=2540003;
RA Farres J., Guan K.L., Weiner H.;
RT "Primary structures of rat and bovine liver mitochondrial aldehyde
RT dehydrogenases deduced from cDNA sequences.";
RL Eur. J. Biochem 180:67-74(1989).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS).
RX MEDLINE=97341232; PubMed=9195888;
RA Steinmetz C.G., Xia P., Weiner H., Hurley T.F.;
RT "Structure of mitochondrial aldehyde dehydrogenase: the genetic

```

RT	component of ethanol aversion.";
RL	Structure 5:701-711(1997).
CC	-1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC	-1- PATHWAY: Ethanol utilization; second step.
CC	-1- SUBUNIT: HOMOTETRAMER
CC	-1- SUBCELLULAR LOCATION: Mitochondrial matrix
CC	-1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
PIR:	S03565; S03565.
PIR:	S09030; S09030.
PDB:	1AG8; 08-OCT-97.
PIR:	1A47; 08-APP-98.
DE	Interf. IFRC02086, Aldehyd-3-hydr.
DR	Pfam: PF00171; aldedh; 1
DR	PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DE	PROSITE: PS00087, ALDEHYDE_DEHYDR_GLU; 1.
KW	oxidoreductase; NAD; Mitochondrion; Transit peptide; 3D structure.
FT	TRANSIT 1 21 MITOCHONDRION.
FT	CHAIN 22 520 ALDEHYDE DEHYDROGENASE.
FT	NIL-IND; 265 270 NAI (ACP-PAM) (BY SIMILARITY).
FT	ACT-SITE 288 288 BY SIMILARITY.
FT	ACT-SITE 322 322 BY SIMILARITY.
SQ	SEQUENCE 530 AA; 56708 MW; P920BCF05E5AE5B CPO64;
Query Match 92.8%; Score 2436; DP 1; Length 520;	
Best Local Similarity 91.8%; Prod. No. 1.4e-179;	
Matches 457, Conservative 21, Mismatches 20, Indels 0, Gaps 0;	
QY	3 AATQAVPAPNQDEVCNGLFINNEWHDAVSKTEPTIVNSTGEVQCVAEGKEDVDK 62
DB	
QY	22 SAATQAVPAPNQDEVLNYGFINNEWHDAVSKTEPTIVNSTGEVQCVAEGKAUVDK 81
DB	
QY	63 APFGPGAPQFGSPWRPMASHSGPLINPLANLTPDPITYLAALFTLNGKPYIISLVLD 122
DB	
QY	82 AVFAAPAAAFQSGSPWPMDSRGPILLNPLADLIERDRTYLAALFTLNGKPYIISLVLD 141
DB	
QY	123 LDMVLKLPYYACWAKYHKTTIPDGLDFSYTHPEPQVQDQIIPWNFPLLMGAWLGP 182
DB	
QY	142 LDMVLKLPYYASWAKYHKTTIPDGLDFSYTHPEPQVQDQIIPWNFPLLMGAWLGP 201
DB	
QY	183 ALATGNVVMKVAEQDTLTALYVANLIKAGFTFGVNLVFGFGTACAAIASHEVDKV 242
DB	
QY	202 ALATGNVVMKVAEQDTLTALYVANLIKAGFTFGVNLVFGFGTACAAIASHEVDKV 261
DB	
QY	243 AFTGSLEICRVQVAAGSSMLKKVYLEGSKSINILMSLDUMDAVGAHIALFNQGC 302
DB	
QY	262 AFTGSLEVGHEIQVAAGSKSNKAVTLEIGSKSFNILMSLDUMDAVGAHIALFNQGC 321
DB	
QY	303 CCASRFTFQVEDIYAEVSVAPAKSPVGNPFTSKTEGPGQVTFQTKKILGYINTGK 362
DB	
QY	322 CCASRFTFQVEDIYAEVSVAPAKSPVGNPFTSKTEGPGQVTFQTKKILGYINTGK 381
DB	
QY	363 QBGAKLLGGGIAADRCGYFTOPTFQVGEVDQNTIAKEELFPGVMOILKFKTIEFVVG 422
DB	
QY	382 PEGILKLTSGGTAAALRGVFLQPTVFELQLQGMILAKELFGVMILKKKSMEEVVG 441
DB	
QY	423 NSGLCLAAAVTKLVNPNVLSLALCATVWVWNYVFGNGSFFFGYVMSGSGPTFGY 482
DB	
QY	442 NSKYGLAAAVTKLVNPNVLSLALGAGVWVNYVWFGVSGSPFGYKLSGSGELG 501
DB	
QY	483 LQAVTEKTVTKVPOKN 500
DB	
QY	502 LQAVTEKTVTKVPOKN 519
DB	
RESULT 6	
DHAM_HORSE	
ID	DHAM_HORSE STANDARD; PRF; 500 AA.
AC	P12762;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-OCT-1989 (Rel. 12, Last sequence update)
DT	15-MAY-2002 (Rel. 41, Last annotation update)
DE	Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3) (AlDH class 2)
DE	(ALDH1) (ALDH-E2).

(BY SIMILARITY).
 - CATALYTIC ACTIVITY: An aldehyde + NAD(+) = an acid + NADH
 - PATHWAY: Ethanol utilization, second step.
 - SUBUNIT: HOMOHEPAMEK
 - SUBCELLULAR LOCATION: Cytoplasmic.
 - SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M31984; AA51692.1; JOINED
 EMBL: M31982; AA51692.1; JOINED
 EMBL: M31983; AA51692.1; JOINED
 EMBL: M31984; AA51692.1; JOINED
 EMBL: M31985; AA51692.1; JOINED
 EMBL: M31986; AA51692.1; JOINED
 EMBL: M31987; AA51692.1; JOINED
 EMBL: M31988; AA51692.1; JOINED
 EMBL: M31989; AA51692.1; JOINED
 EMBL: M31990; AA51692.1; JOINED
 EMBL: M31991; AA51692.1; JOINED
 EMBL: M31992; AA51692.1; JOINED
 EMBL: AF003341; AC051652.1;
 EMBL: K03000; AA51695.1;
 EMBL: M26761; AA35519.1;
 EMBL: S61345; A013925.1;
 PIR: A33371; DEH01.
 HSSP: P51977; 1BXS.
 SWISS-2DPAGE: P00352; HUMAN.
 Genew: HGNC:402; ALDH1A1.
 MIM: 100640;
 InterPro: IPR002086; Aldehyde_d-hydr
 Pfam: PF00171; aldehyd; 1.
 PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1
 PROSITE: PS00687; ALDEHYDE_DEHYDR_GLIF; 1.
 KW: oxidoreductase; NAD; Acetylation
 INIT_MET 0
 MOD_RES 1 1 ACETYLATION
 NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
 ACT_SITE 248 268 POTENTIAL.
 ACT_SITE 302 302 POTENTIAL.
 CONFLICT 120 120 N -> S (IN REF. 2).
 CONFLICT 161 161 V -> I (IN REF. 3 AND 4).
 SEQUENCE 500 AA; 54730 MW; 8264541454749677 CP; 64;

Query Match 67.4%; Score 1769; DB 1; Length 500;
 Best Local Similarity 66.9%; Pred. No. 3.3e-128;
 Matches 334; Conservative 67; Mismatches 98; Indels 0; Gaps 0.
 QY 2 SAAATGAVFAINQCIIEVCNO;FINNEWHUAVSKKFFIVNLSIGEVILAVAGDIOKDW;61
 Db 1 SSSGTPLDVLTLTKIYKIPFINNEWHUAVSKKFFIVNLSIGEVILAVAGDIOKDW;60
 QY 62 KAREKRGAPVAGSPPWRMIAASHKPLNPLAMIFKQFTYIAALETLNENKPYVLSLV 121
 Db 61 KAVKAAACAPGTSPPWTPDASEPGPIIYKIDIFEDPILIAATMESMNGSKLYSNAYLN 120
 QY 122 QLDMLVCLRYACWAKYHKTIFDSTFSTYHFPVGVQGITFWNPLIMQAWKLG 181
 Db 121 QLDMLVCLRYACWAKYHKTIFDSTFSTYHFPVGVQGITFWNPLIMQAWKLG 180
 QY 182 PALATGNVVMKVAEOTPLTALYVANLIKEAGPPGVVNIIVPGFGPTAGAAIASHRDVDK 241
 Db 181 PALSCGNVVMVVKPAEOTPLTALYVANLIKEAGPPGVVNIIVPGFGPTAGAAIASHRDVDK 240
 QY 242 VAPTSSTIEIPVIVAAVSSNLKRVTLFEDGKSFNLIIMSDAMWAVESAHFALFTNQCC 301
 Db 241 VAPTSSTIEIPVIVAAVSSNLKRVTLFEDGKSFNLIIMSDAMWAVESAHFALFTNQCC 300

QY 402 QYAGSKTFFVQDEIYDFEFVVSAPAKSPVVNPHFUSKFFGSPVDFLQFKKLLGVINIG 361
 Db 401 QYAGSKTFFVQDEIYDFEFVVSAPAKSPVVNPHFUSKFFGSPVDFLQFKKLLGVINIG 360
 QY 362 KQEKAKLEGGSGIAADGQYFQIPVFCWQFPMTEIAKEETPSIVMAJLKFRTFFVVGRA 421
 Db 361 KQEKAKLEGGSGIAADGQYFQIPVFCWQFPMTEIAKEETPSIVMAJLKFRTFFVVGRA 420
 QY 422 NNSTYGLAAAVFTKCFKANYLSQALAGVWVWVYVDFPAAJSPFQYKMSKSGPFLRY 481
 Db 421 NNSTYGLAAAVFTKCFKANYLSQALAGVWVWVYVDFPAAJSPFQYKMSKSGPFLRY 480
 QY 482 GLQAVTEVKTIVKVPKN 500
 Db 481 GFHEYTEVKTIVTKISOKN 499

RESULT 14

DHAL_SHEEP
 ID DHAL_SHEEP STANDARD; PRT; 500 AA.
 AC P51977;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase 1A1 (F1.1.2.1.3) (Aldehyde dehydrogenase, cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).
 GN ALDH1A1 OR ALDH1.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Primates; Ruminantia; Bovidae; Caprinae; Ovis.
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID:9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 FX MEDLINE=96062535; PubMed=7484119;
 RA Stayner C.K., Tweedie J.W.;
 PT "Cloning and characterization of the cDNA for sheep liver cytosolic aldehyde dehydrogenase."
 PL Adv. Exp. Med. Biol. 372:61-66(1995).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS).
 RC TISSUE=Liver;
 FX MEDLINE=95081884; PubMed=9852807;
 RA Moore S.A., Baker H.M., Rytche T.J., Kitson K.E., Kitson T.M., Baker E.N.;
 PT "Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases."
 PL Structure 6:1541-1551(1998).
 CC - FUNCTION: BINDS FREE RETINAL AND CELLULAR RETINOL; PROTEIN-BINDING RETINAL. CAN CONVERT XENOBOTIC ALDEHYDES TO RETINOIC ACID.
 CC - CATALYTIC ACTIVITY: An aldehyde + NAD(+) = an acid + NADH.
 CC - PATHWAY: Ethanol utilization, second step.
 CC - SUBUNIT: HOMOHEPAMEK.
 CC - SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U12761; AAA85435.1;
 DB: LUXS; 27-APR-99.
 DB InterPro: IPR002086; Aldehyde_dehydratase; 1.
 DB PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1.
 DB PROSITE: PS00687; ALDEHYDE_DEHYDR_GLIF; 1.

```

KW Oxidoreductase; NAD, Acetylation; 3D-structure.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268
FT ACT_SITE 302 302
SQ SEQUENCE 500 AA; 54693 MW; 68B8971968CEB8F8 CRC64;

Query Match 67.0%; Score 1758; DB 1; Length 500;
Best Local Similarity 66.1%; Pred. No. 2,3e-127;
Matches 332; Conservative 67; Mismatches 100; Indels 0; Gaps 0;

QY 2 SAATAQVAPAPNOQPEVFCNQIFINNEHDAVSRRKIFTPVNPSTGEVICVAGDKEDVD 61
DB 1 SSSAMPDVPAPLTNIQFKYTKIFINNEHSSVSGKKFPVNPATEEKLCEVEBGDKEDVD 60
QY 62 KAREGRGAFQIGSPWRMDASHSGRLNRLADLIFRDRTYLAALETLDNGKPKYVSYLV 121
DB 1 KAVKAARQAFQIGSPWRMDASHSGRLNRLADLIFRDRTYLAALETLDNGKPKYVSYLV 121
QY 61 KAVKAARQAFQIGSPWRMDASHSGRLNRLADLIFRDRTYLAALETLDNGKPKYVSYLV 121
DB 1 KAVKAARQAFQIGSPWRMDASHSGRLNRLADLIFRDRTYLAALETLDNGKPKYVSYLV 121
QY 122 DLDMVLKCLPYPAWAPYHCKTIPIDGDFPSYTRHPVGVCGOIIIPNFPPLMOAWKLG 181
DB 122 DLDMVLKCLPYPAWAPYHCKTIPIDGDFPSYTRHPVGVCGOIIIPNFPPLMOAWKLG 181
QY 121 DLGGCIKILYACAWAKLQCKRIIPMGNFFIYKSEPVGVCGQIIPNFPPLMOAWKLG 180
DB 121 DLGGCIKILYACAWAKLQCKRIIPMGNFFIYKSEPVGVCGQIIPNFPPLMOAWKLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPVGGPTAGAAIASHEDYDK 241
DB 182 PALATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPVGGPTAGAAIASHEDYDK 241
QY 181 FALSCGNTVVKFAEOTPLTALHMSGLIKEAGFPFGVNVNIPVGGPTAGAAIASHEDYDK 240
DB 181 FALSCGNTVVKFAEOTPLTALHMSGLIKEAGFPFGVNVNIPVGGPTAGAAIASHEDYDK 240
QY 242 VAFGSTEICRVLOVAGSSNLIKRVTLGLGKSPNIMSDAIDMWAVEQAHFALFPNGQ 301
DB 242 VAFGSTEICRVLOVAGSSNLIKRVTLGLGKSPNIMSDAIDMWAVEQAHFALFPNGQ 301
QY 241 VAFGSTEICRVLOVAGSSNLIKRVTLGLGKSPNIMSDAIDMWAVEQAHFALFPNGQ 300
DB 241 VAFGSTEICRVLOVAGSSNLIKRVTLGLGKSPNIMSDAIDMWAVEQAHFALFPNGQ 300
QY 302 CCAGSKIEVQEDIDYDFVRSVARAKSRVGVNPFDSKTEQGPVDFQPKKILGYNTG 361
DB 302 CCAGSKIEVQEDIDYDFVRSVARAKSRVGVNPFDSKTEQGPVDFQPKKILGYNTG 361
QY 301 CCIAASKIEVQEDIDYDFVRSVARAKSRVGVNPFDSKTEQGPVDFQPKKILGYNTG 360
DB 301 CCIAASKIEVQEDIDYDFVRSVARAKSRVGVNPFDSKTEQGPVDFQPKKILGYNTG 360
QY 362 KQEGAKLLGGGIAADKGYFIQPTVFGVDQGMTIAKEIEIPGVNQILKFKTIEEVVGRA 421
DB 362 KQEGAKLLGGGIAADKGYFIQPTVFGVDQGMTIAKEIEIPGVNQILKFKTIEEVVGRA 421
QY 361 KREGAKLEGCGGPGWNGVFIQPTVFGVDQGMTIAKEIEIPGVNQILKFKTIEEVVGRA 420
DB 361 KREGAKLEGCGGPGWNGVFIQPTVFGVDQGMTIAKEIEIPGVNQILKFKTIEEVVGRA 420
QY 422 NNTYGLAAVFTKDLKANYLSOALQACTVWVNCYDFVFAQSPGPKMSSSKPLGEY 481
DB 422 NNTYGLAAVFTKDLKANYLSOALQACTVWVNCYDFVFAQSPGPKMSSSKPLGEY 481
QY 421 NNTYGLSAGFTINDIKAITVSSALQSGTVWVNCYSVVSACCPGPGFKMSGNGRELGEY 480
DB 421 NNTYGLSAGFTINDIKAITVSSALQSGTVWVNCYSVVSACCPGPGFKMSGNGRELGEY 480
QY 482 GLQAYTEKTVTKVVKPKN 500
DB 482 GLQAYTEKTVTKVVKPKN 500
QY 481 GPHEYTEKTVTKIKISOKN 499
DB 481 GPHEYTEKTVTKIKISOKN 499

RESULT 15
DHAI_MOUSE STANDARD; PRT; 500 AA.
AC P24549;
DT 01-MAR-1992 (Rel. 21, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (EC 1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).
OS ALDH1A1 OR ALDH1 OR ALDH2 OR ALDH3
CN Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
PC STPAIN=BALB/c, and C57BL/6; TISSUE=Liver;
FX MEDLINE=91276281; PubMed=2055490;
RA Rongoparut P. Weaver S.;
RT "Isolation and characterization of a cytosolic aldehyde
RT dehydrogenase-encoding cDNA from mouse liver.";
KL Gene 101:261-265(1991).
RN [2]
RP SEQUENCE FROM N.A.

```



```

Db      181 PALSOGNTVVVKPAFQPTLTAHLASLIKEAGFPPTVVNTVPGYGPPTAGAAISSHMDVVK 240
QY      242 VAFTGSTTIGRVIQVAAGSSNLKRVTLIELGSKSPNTIMSDADMWAVEQAHFALFENQJQ 301
Db      241 VAFTGSTGVGKLIKEAAGKSNLKRVTILELGGKSPCIVFADADLDIAVEFAHGVYHQJQ 300
QY      302 CTCAGSPAFVQEDIVDEFVHVSAPAKSPVWGNPFUSKTEQGGCVGDEIQFKKILGYINTG 361
Db      301 CCVAASRIFVESYDEVRKSVVERAKKYVLGNPLTPGINQGPQIDKEQHDKILDLESG 360
QY      362 KQERAKLIGRHSIAAPRPFYFIQPIVPHWJFGMTIAKEELFGPVWQILKPKTILEEVVQPA 421
Db      361 KKEGAKLECGGRFNGKGFVQPIVFSNVITLMEFLAKEELFGPVQQLMKFKSVDDVIKPA 420
QY      422 NNSTVGLAAAVFTKDLKANYLSQALGAGTVWVNTVIVFGAGSPFGGKMSGSGRELGEY 481
Db      421 NNTYGLAAGLTKDLDRATVSSALQAGVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVW 480
QY      482 GLQAYTEVKTVTYKVPQKN 500
Db      481 GLYEYTELKTVMKISQKN 499

```

Search completed. June 24, 2003, 10 14 28
 Job time : 12 0705 secs




```

QY 179 KLGPALATGNVVMKVAHPTPLTALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 238
DB 121 KLGPALATGNVVMKVAHPTPLTALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 180
QY 239 VDKVAFVTSIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMDWAVEQAHPALFFN 298
DB 181 VDKVAFVTSIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMDWAVEQAHPALFFN 240
QY 299 QGCCAGSRTFVOEDIDYDFVVRVSARAKSVVGNPFDSKTEQGPQVDETQFKKILGYI 358
DB 241 QGCCAGSRTFVOEDIDYDFVVRVSARAKSVVGNPFDSKTEQGPQVDETQFKKILGYI 300
QY 359 NTGKQEGAKLCCGGIAADRGYFIOTVFGDVODGMT:AKEEIFGPMVQIILKFKTIERVV 418
DB 301 KSGQEGAKLCCGGIAADRGYFIOTVFGDVODGMT:AKEEIFGPMVQIILKFKTIERVV 360
QY 419 GRANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDFVFGAGSPGQYKMSGSGREL 478
DB 361 GRANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDFVFGAGSPGQYKMSGSGREL 420
QY 479 GEYGLQAYTEVKTIVTKVPQKN 500
DB 421 GEYGLQAYTEVKTIVTKVPQKN 442

RESULT 2
O8QGQ2 PRELIMINARY: PRT: 516 AA.
AC O8QGQ2:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Aldehyde dehydrogenase 2.
GN ALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Taniguchi R.L., Vasilou V.:
RT "Expression and characterization of mitochondrial aldehyde
RL dehydrogenase (ALDH2) in zebrafish (Danio rerio).";
RL submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF260121; AAM19352.1;
SQ SEQUENCE 516 AA; 56562 MW; EA7D873F66905821 CPr64;

Query Match 74.9%; Score 2097; DB 13; Length 516;
Best Local Similarity 77.8%; Pred No 1 7p-161;
Matches 389; Conservative 53; Mismatches 58; Indels 0; Gaps 0;

QY 1 MSAATQAVAPNPOOPEVCNOIFINNEWHDAVSRKTEPTVNPSTGEVICOVARGKEDV 60
DB 16 ISSQHSHTIPAPNVQPDVHYNKIFINNEWHDAVSRKTEPTINPATAEVICVARGKADV 75
QY 61 DKAREGRPGAFQIGSPWRMDASHGRLNRLADLIERDRTYLALETDLNGKPYVLSYL 120
DB 76 DKAVKAARDAPFKIGSHWRMPDASQGLLLNPLAOTEPAAAYLAFTLTLNGKPYTISPT 135
QY 121 VLDLWVLKCLRYAGWADKYHCKTIPIDGDFSFSTRPEPVVCCQIIPWNPPLLMQAWKL 180
DB 136 VDLPMVVKCLRYAGWADKWCKTIPIDGNVFCYTRIEPICVCGQIIPWNPPLLMQAWKL 195
QY 181 GPALATGNVVMKVAHPTPLTALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 240
DB 196 GPALATGNVVMKVAHPTPLTALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 255
QY 241 KVAFVTSIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMDWAVEQAHPALFFNQ 300
DB 256 KVAFVTSIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMDWAVEQAHPALFFNQ 315

```

```

QY 301 QGCCAGSRTFVOEDIDYDFVVRVSARAKSVVGNPFDSKTEQGPQVDETQFKKILGYINI 460
DB 316 QGCCAGSRTFVOEDIDYDFVVRVSARAKSVVGNPFDSKTEQGPQVDETQFKKILGYISS 475
QY 361 GKQEGAKLCCGGIAADRGYFIOTVFGDVODGMT:AKEEIFGPMVQIILKFKTIERVVGR 420
DB 376 GKQEGAKLCCGGIAADRGYFIOTVFGDVODGMT:AKEEIFGPMVQIILKFKTIERVVGR 445
QY 421 ANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDFVFGAGSPGQYKMSGSGREL 480
DB 436 ANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDFVFGAGSPGQYKMSGSGREL 495
QY 481 YGLQAYTEVKTIVTKVPQKN 500
DB 496 YGLQAYTEVKTIVTKVPQKN 515

RESULT 3
Q9CZSL PRELIMINARY: PRT: 519 AA.
AC Q9CZSL:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 31 JUN 2002 (TREMBLrel. 21, Last annotation update)
DE 2700007F14Rik protein (Aldehyde dehydrogenase 1 family, member
DE B1).
GN ALDH1B1 OR 2700007F14Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRIO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata T.,
RA Saito I., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleisemann W., Gaasterland T., Gissi C., King R., Korchova H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakai T., Pesole G., Quackenbush J.,
RA Sakai K., Staubli F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anno H., Haidarali R., Harsh G.,
RA Blake J., Boffelli D., Bojarski N., Catinelli P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustineich S., Hill D., Holman M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombertis P.,
RA Nordone P., Ring H., Ringwald M., Rodriguez T., Sakamoto N.,
RA Suzuki H., Sato K., Schenck C., Seya T., Shibata Y., Storch K.,
RA Wynshaw-Boris A., Yoshida K., Wang R.H., Wolfe C., Whittaker C., Wilming L.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AK01211; BAB28101.1;
DR EMBL: BC020001; AAB20001.1;
DR HSP: P05091; LCW3.
DR MCD: MGI:191785; Aldh1b1.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
DR Oxidoreductase.
SQ SEQUENCE 519 AA; 57552 MW; 6322217409BEEF3 CPr64;

Query Match 74.5%; Score 1956; DB 11; Length 519;

```


Db 138 DLDEVIVKRYFASWAKKHGKTIIPMDGQHCFTRHEPVGVCQIIPWNPFLVMGCKMLA 197
 QY 182 PALATGNNVVMKVAEOTPLTALYVANLKEAGFPFGVNVIVPGFPTAGAAATASHEDYDK 241
 Db 198 PALATGNNVVMKVAEOTPLTALYVANLKEAGFPFGVNVIVPGFPTAGAAATASHEDYDK 257
 QY 242 VAFSTGEIGRVIVQAAACSNLKRVTLELCKGSPNIDMSDADMDWAVEGAIFALFFNGQ 301
 Db 258 VAFSTGEIGRVIVQAAACSNLKRVTLELCKGSPNIDMSDADMDWAVEGAIFALFFNGQ 317
 QY 302 CCAGSGTFFVOEDYDEFVRSVARAKSRVGNCPNDSKTEGQPOVDOTFKKILCYLNG 361
 Db 318 CCAGSGTFFVOEDYDEFVRSVARAKSRVGNCPNDSKTEGQPOVDOTFKKILCYLNG 377
 QY 362 KQRCAGLIGGTAANDGYFTQPTVPBIVQVQMTIAKEIEFGVNMQILKFKTIEVVGRA 421
 Db 378 QKRGKALICGRRFGGRRGFFIKPTVFGVQVQVMPIAKEIEFGVNMQILKFKTIEVVGRA 437
 QY 422 NSTSYGLAAAVFTKDLKKNYLSQLAQAGTVVWNTYVFGAUSPRGYSKMSGSGRELGEY 481
 Db 438 NTRYGLAAAVFTKDLKKNYLSQLAQAGTVVWNTYVFGAUSPRGYSKMSGSGRELGEY 497
 QY 482 GLQAYTEVKTVIVKVPQKN 500
 Db 498 GLKAYTEVKTVIVKVPQKN 516

RESULT 6

QYVLC5
 ID Q9VLC5 PRELIMINARY, PRT, 520 AA.
 AC Q9VLC5;
 DT 01-MAY-2000 (TREMblrel, 13, created)
 DT 01-MAY-2000 (TREMblrel, 13, last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, last annotation update)
 DE CG3752 protein.
 GN CG3752.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta;
 OC Pterygota, Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
 OC Ephydroidea, Drosophilidae, Drosophila.
 OX NCBI_TaxID=7227;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 FX MEDLINE=20196066; PubMed=10731132;
 RA Adams M.D., Galinier S.E., Holt P.A., Evans C.A., Gwynne D.B.
 RA Ananadides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon P., Rogers Y.-H., Blazej P.G., Chapple M., Pfeiffer R.P.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berland J.P., Blandat D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Curtis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Finkler C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris T., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy K., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.P., Paulep J.M.,
 RA Palazzi G.M., Pittman G.S., Pan S., Pollard T., Puri V., Reese M.S.,

RA Reinert K., Remington K., Saunders K.D.C., Schechter F., Shen B.,
 RA Shue H.C., Slonikamos L., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector R., Turner R., Venter E., Ward A.H., Ward X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach L.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh K.F., Zayori J.S., Zhao M., Zhang G., Zhao Q., Zhou L.,
 RA Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith R.G.,
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL: AF003624; AAF52769.1; ...
 DR HSP: P05091; ICW3.
 DR FlyBase: Fbm0032114; CG1752.
 DR InterPro: IPR002086; Ald-aldh-dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR. CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR. GID; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 520 AA; 57019 MW; 82HF54D573EC77B CR664;
 Query Match 71.1%; Score 1866; DR 5; Length 520;
 Best Local Similarity 69.8%; Pred. No. 9, 3e-14;
 Matches 351; Conservative 61; Mismatches 84; Indels 8; Gaps 4;
 QY 2 SAAATQAVAPNCOPEVFCNQIFINNEWHDAVSRKTPPTVNPSTGEVLCVAGKELVD 61
 Db 21 AVANYSSLPQPTTPTDLYTVTFINNHKRSKGFETINTTAETVIAEQCAKREID 80
 QY 62 KAREGRTAFQAGSGPWKMIASHSGRLKLAHLERKTYLAALFELNKKPVVSYLV 121
 Db 81 IAVQAAPNAFKLGSPWPMASEPGLYPLALIMEPLAVVLALEFTEPKPKYSNV 140
 QY 122 ELIMVLELLEPYTACWALVYDLELLELLELLELLELLELLELLELLELLELLEL 181
 Db 141 LELTALKELEYFAGWAGNKELELELELELELELELELELELELELELELELE 200
 QY 182 PALATGNNVVMKVAEOTPLTALYVANLKEAGFPFGVNVIVPGFPTAGAAATASHEDYDK 241
 Db 201 PALATGNNVVMKVAEOTPLTALYVANLKEAGFPFGVNVIVPGFPTAGAAATASHEDYDK 259
 QY 242 VAFSTGEIGRVIVQAAACSNLKRVTLELCKGSPNIDMSDADMDWAVEGAIFALFFNGQ 301
 Db 260 VAFSTGEIGRVIVQAAACSNLKRVTLELCKGSPNIDMSDADMDWAVEGAIFALFFNGQ 319
 QY 302 CCAGSGTFFVOEDYDEFVRSVARAKSRVGNCPNDSKTEGQPOVDOTFKKILCYLNG 361
 Db 320 CCAGSGTFFVOEDYDEFVRSVARAKSRVGNCPNDSKTEGQPOVDOTFKKILCYLNG 379
 QY 362 KQRCAGLIGGTAANDGYFTQPTVPBIVQVQMTIAKEIEFGVNMQILKFKTIEVVGRA 421
 Db 380 KQRCAGLIGGTAANDGYFTQPTVPBIVQVQMTIAKEIEFGVNMQILKFKTIEVVGRA 436
 QY 418 VAFSTGEIGRVIVQAAACSNLKRVTLELCKGSPNIDMSDADMDWAVEGAIFALFFNGQ 477
 Db 437 IFRANNSEYGLAAAVFTKDLKKNYLSQLAQAGTVVWNTYVFGAUSPRGYSKMSGSGRELGEY 496
 QY 478 LGEYGLQAYTEVKTVIVKVPQKN 500
 Db 497 NCEYALSNTEVKSIVIVKVALKN 519
 ID Q9VLC5 PRELIMINARY, PRT, 518 AA.
 AC Q9VLC5;
 DT 01-MAR-2001 (TREMblrel, 16, created)
 DT 01-MAR-2001 (TREMblrel, 16, last sequence update)
 DT 01-MAR-2002 (TREMblrel, 21, last annotation update)
 DE RALDH2.
 OS Xenopus laevis (African clawed frog).
 OX NCBI_TaxID=7227;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 FX MEDLINE=20196066; PubMed=10731132;
 RA Adams M.D., Galinier S.E., Holt P.A., Evans C.A., Gwynne D.B.
 RA Ananadides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon P., Rogers Y.-H., Blazej P.G., Chapple M., Pfeiffer R.P.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berland J.P., Blandat D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Curtis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Finkler C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris T., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy K., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.P., Paulep J.M.,
 RA Palazzi G.M., Pittman G.S., Pan S., Pollard T., Puri V., Reese M.S.,

Aldehyde dehydrogenase (EC 1.2.1.3).

GN ALDH1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95191703; PubMed=10091603;
RA Ang H.L., Duester G.;
RT "Stimulation of premature retinoic acid synthesis in Xenopus embryos
following premature expression of aldehyde dehydrogenase ALDH1.";
RL Eur. J. Biochem. 260:237-244(1999).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR FMIL: AF061833; AAC69552.1;
DR HSP: P51977; IBSX.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 502 AA; 55205 MW; 54646 Thr56; AA710; C6264.

Query Match 68.3%; Score 1793; DB 13; Length 502;
Best Local Similarity 67.1%; Pred. No. 7.4e-137;
Matches 330; Conservative 67; Mismatches 95; Indels 0; Gaps 0;

QY 9 VPAPNOQPEVFCNQIFINNEWHDAVSRKTFPTVNVSTGEVTCQVAGCKEDVDKAREGR 68
DB 10 LPPLSLNLEIKYTKIFINNEWINSISGKKFPYNNATGEKVEEGKDFVKNVAKAAR 69
QY 69 GAFQGLSPWRPMASHSGRLNKLAFLEPDTLAALEFLINRKPVVISLVLIQLMVLR 129
DB 70 FAEQGLSPWRPLDASERGMLNKLADLVERDPLILSTWESINCSKPYTASTGEIFCAIK 129
QY 129 CLRYAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLMLQAWKLGPAATGN 188
DB 130 SLRYCAWADKVGRTIPMGAGYFTTRHEPVGVGGQIIPWNPFLVMEAWKIAVALCGN 189
QY 189 VVYMKVAEQTPITALYVAVNLKEAGFPFGVNNIPVPGPTAGAAATASHEDVDKVAFTGST 248
DB 190 TVYMKPAEQTPITALYVAVNLKEAGIPFGVNNIPVPGPTAGAAATASHEDVDKVAFTGST 249
QY 249 EIGRPIQVAAAGSSNLKPVTLLEGGKSPNITMSDADMWAVECAHFALEFALFPTNQGGV 308
DB 250 EVCKLKEAAGKSNLKVTLLEGGKSPNITFADALLDAVEHAHNLFFHHQPCNLAGSR 309
QY 309 TFVQEDLYELFVYKSVARAKSVVGNPFDCKTEGHPQVDETQPKLLGYINTGKQGGAKL 368
DB 310 IPVEEPIYDEFVRSVERAKKRVLDGDPAPVYNDGQGLCKEYQKCLELIESGKKEGAKL 369
QY 369 LQSGTIAAGRGYFIQPTVFGVQGMFTAKEEIPGVVMTLLKFTIEEVGSPANNSTYGL 428
DB 370 QCGGSAMGEGYFISTVFSVDKDDMR TAKEELFPGVQVLLKFTIIDEIVKRNATKYGL 429
QY 429 AAATVTKDLKANYLSOALQATVWVNYDVGAGSPFGYKMSGSGHEIGYGLQVATE 489
DB 430 AAGVFTKDMKATILMSTALOAGTVWVNYDVGAGSPFGYKMSGSGHEIGYGLQVATE 489
QY 489 VKTIVTVKVPQKN 500
DB 490 VKTIVTKISQKN 501

RESULT 10
Q90XSR
ID Q90XSR PRELIMINARY; PPT: 518 AA.
AC Q90XSR;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Retinaldehyde dehydrogenase type 2.

GN RALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21543521; PubMed 11488568;
RA Bequeman G., Schilling T.F., Rauch G., J., Geisler K., Ingham P.W.;
RT "The zebrafish neofunctional mutation reveals a requirement for RALDH2 in
mesodermal signals that pattern the hindbrain.";
KL Development 128:3081-3094(2001).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMIL: AF348437; AAL00899.1;
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNWN.1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNWN.2.
KW Oxidoreductase.
SQ SEQUENCE 518 AA; 56515 MW; 2060451617E8D86; C6064.

Query Match 68.0%; Score 1786; DB 14; Length 518;
Best Local Similarity 66.9%; Pred. No. 2.9e-140;
Matches 332; Conservative 65; Mismatches 99; Indels 0; Gaps 0;

QY 4 AATQAVPAPEVFCNQIFINNEWHDAVSRKTFPTVNVSTGEVTCQVAGCKEDVDKAREGR 64
DB 21 ASJHLPSPVPEPTKYTFINNEWHDSVSKVETTYNDATGKPTIVGFAKADVUKA 80
QY 64 KEGPGAGFAGSSNKPWPMASHSGRLNKLAFLEPDTLAALEFLINRKPVVISLVLIQL 124
DB 81 VQAAPEASGLASVWPMASHSGRLNKLAFLEPDTLAALEFLINRKPVVISLVLIQL 140
QY 124 DMVLCRLRYAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLMLQAWKLGPA 184
DB 141 OGILKTFRYAGWADKIHGSTIPIDGDFSYTRHEPVGVGGQIIPWNPFLVMTAKLGPA 200
QY 184 LATGNVVMKVAEQTPITALYVAVNLKEAGFPFGVNNIPVPGPTAGAAATASHEDVDKVA 244
DB 201 LSGNIVLVKPAEQTPITALYVAVNLKEAGFPFGVNNIPVPGPTAGAAATASHEDVDKVA 260
QY 244 FTGSTEIGRPIQVAAAGSSNLKPVTLLEGGKSPNITMSDADMWAVECAHFALEFALFPTNQGGV 304
DB 261 FTGSTEIGRPIQVAAAGSSNLKPVTLLEGGKSPNITMSDADMWAVECAHFALEFALFPTNQGGV 320
QY 304 CAGSPTEVQEDLYELFVYKSVARAKSVVGNPFDCKTEGHPQVDETQPKLLGYINTGKQ 364
DB 321 TAGSRIFVEEPIYDEFVRSVERAKKRVLDGDPAPVYNDGQGLCKEYQKCLELIESGKKEGAKL 380
QY 364 EGAKLGGGIAAGRGYFIQPTVFGVQGMFTAKEEIPGVVMTLLKFTIEEVGSPANN 424
DB 381 EGAKLGGGIAAGRGYFIQPTVFGVQGMFTAKEEIPGVVMTLLKFTIEEVGSPANN 440
QY 424 STYGLAAATVTKDLKANYLSOALQATVWVNYDVGAGSPFGYKMSGSGHEIGYGL 484
DB 441 STYGLAAATVTKDLKANYLSOALQATVWVNYDVGAGSPFGYKMSGSGHEIGYGL 500
QY 484 QAVTEVTVKVPQKN 499
DB 501 KEYTELKTIITMKMSCK 516

RESULT 11
Q8QGO1
ID Q8QGO1 PRELIMINARY; PPT: 518 AA.
AC Q8QGO1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Retinaldehyde dehydrogenase 2.
GN RALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).

OC Eukaryota, Metazoa; Chordata, Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes.
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Grandel H., Lin K., Pauch G.-J., Phinn M., Piotrowski T., Hewart C.,
 RA Sordino P., Kuechler A.M., Schulte-Merker S., Geisler R., Holder N.,
 RA Wilson S.W., Brand M.;
 RT "Retinoic acid signalling in the zebrafish embryo is necessary during
 RT pre-segmentation stages to pattern the anterior-posterior axis of the
 RT CNS and to induce a pectoral fin bud.";
 RT Development 0.0.0(2002).
 DR EMBL: AF288764; AAK83071.2; -;
 SQ SEQUENCE 518 AA: 56579 MW: 406777974EA CPC64;

Query Match 67.7%; Score 1777; DB 13; Length 518;
 Best Local Similarity 66.7%; Pred. No. 3.9e-135;
 Matches 331, Conservative 65, Mismatches 100, Indels 0, Gaps 0.
 QY 4 AATQAVPAPNQPEVFCNQIFINNEHDAVSRTFTPTVPSTGEVICOVARGKUDVKA 63
 DB 21 ASLHLMSPVNPKEIKYIFINNEHDSVSGVFTYNPATGEKICDVOEADKADVDKA 80
 QY 64 PEPFPAFQSGSPWPMATASHSGPILNPLADLIERDPTLYLALETLDNGKPYVIVYL 123
 DB 81 VAAKSAFSLGVSVMKMUASEKGLKLLADLVVERDSAYLALLESLSGKPLPCFFVUL 140
 QY 124 DWLKLRLRYAGWADKYHGKTIPIIDGPFSSYTPHPFVGVCGQIIPNFFLLMOAKLGP 183
 DB 141 QGIIKTRFYAGWADKIRGSTIPIIDGFEFTLRHEPIGVCGQIIPNFFLVMTANKLGP 200
 QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPGFGPTAGAAIASHEDVDKA 243
 DB 201 LSCGNTVVLKPAEOTPLTCLYLGALKEAGFPFGVNVNIPGFGPTAGAAISSHMGIDKA 260
 QY 244 FTGSTEGRVQVAAAGSNLKRVTLELGGKSPNII MSDADMVAQAHFALFFNQGCC 303
 DB 261 FTGTEGKLVQAEAGSNLKRVTLELGGKSPNII IFADADFLALBQAHQGVFFNQGCC 320
 QY 304 CAGSRTFVQEDIDYDFVVRVSARAKSVNPNPDSKTEQGPVDFTQPKILGYNTGK 363
 DB 321 TAGSPIFVEEPIYDFVVRVSARAKSVNPNPDSKTEQGPVDFTQPKILGYNTGK 380
 QY 364 EGAKLGGGIIAAGVYFIQPTVFGVQDGMTAKFFIFGPMQIILKPTTFFVVPANN 423
 DB 381 EGAKLEGCGKAPATKGEFVEPTVFSNVKDHMRKAKEEIEGPGVQIIMKFTIEVIERANN 440
 QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVWVNYGVFGAGSPGGYKMSGSKELCEYGL 483
 DB 441 TEYGLAAAVFTDTSKAMTISAAVQAGTVWVNYGVFGAGSPGGYKMSGSKELCEYGL 500
 QY 484 QAYTEVKTVTKVPQK 499
 DB 501 KEYTELKTIIMKMSGK 516

RESULT 12
 Q09Y03
 ID Q09Y03 PRELIMINARY; PRT; 518 AA.
 AC Q09Y03
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE Aldehyde dehydrogenase 1A2.
 DE Brachydanio rerio (zebrafish).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tanguay R.L., Pappa A., Vasilion V.;

"Characterization of the zebrafish aldehyde dehydrogenase 1A2.";
 KL Submitted (000000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL: AF315631; AA126332.1; -;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_2.
 KW Oxidoreductase.
 SQ SEQUENCE 518 AA: 56579 MW: 40680648P9A2R CPC64;

Query Match 67.5%; Score 1772; DB 13; Length 518;
 Best Local Similarity 66.5%; Pred. No. 3.9e-135;
 Matches 330, Conservative 65, Mismatches 101, Indels 0, Gaps 0.
 QY 4 AATQAVPAPNQPEVFCNQIFINNEHDAVSRTFTPTVPSTGEVICOVARGKUDVKA 63
 DB 21 ASLHLMSPVNPKEIKYIFINNEHDSVSGVFTYNPATGEKICDVOEADKADVDKA 80
 QY 64 PEPFPAFQSGSPWPMATASHSGPILNPLADLIERDPTLYLALETLDNGKPYVIVYL 123
 DB 81 VAAKSAFSLGVSVMKMUASEKGLKLLADLVVERDSAYLALLESLSGKPLPCFFVUL 140
 QY 124 DWLKLRLRYAGWADKYHGKTIPIIDGPFSSYTPHPFVGVCGQIIPNFFLLMOAKLGP 183
 DB 141 QGIIKTRFYAGWADKIRGSTIPIIDGFEFTLRHEPIGVCGQIIPNFFLVMTANKLGP 200
 QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPGFGPTAGAAIASHEDVDKA 243
 DB 201 LSCGNTVVLKPAEOTPLTCLYLGALKEAGFPFGVNVNIPGFGPTAGAAISSHMGIDKA 260
 QY 244 FTGSTEGRVQVAAAGSNLKRVTLELGGKSPNII MSDADMVAQAHFALFFNQGCC 303
 DB 261 FTGTEGKLVQAEAGSNLKRVTLELGGKSPNII IFADADFLALBQAHQGVFFNQGCC 320
 QY 304 CAGSRTFVQEDIDYDFVVRVSARAKSVNPNPDSKTEQGPVDFTQPKILGYNTGK 363
 DB 321 TAGSPIFVEEPIYDFVVRVSARAKSVNPNPDSKTEQGPVDFTQPKILGYNTGK 380
 QY 364 EGAKLGGGIIAAGVYFIQPTVFGVQDGMTAKFFIFGPMQIILKPTTFFVVPANN 423
 DB 381 EGAKLEGCGKAPATKGEFVEPTVFSNVKDHMRKAKEEIEGPGVQIIMKFTIEVIERANN 440
 QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVWVNYGVFGAGSPGGYKMSGSKELCEYGL 483
 DB 441 TEYGLAAAVFTDTSKAMTISAAVQAGTVWVNYGVFGAGSPGGYKMSGSKELCEYGL 500
 QY 484 QAYTEVKTVTKVPQK 499
 DB 501 KEYTELKTIIMKMSGK 516

RESULT 13
 Q020780
 ID Q020780 PRELIMINARY; PRT; 510 AA.
 AC Q020780
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DE Hypothetical 55.1 kDa protein.
 DE F54D8.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhaditida; Rhaditidae.
 OC Rhaditidae; Pelodidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).

Search completed: June 24, 2003, 10:17:09
Job time : 39.2679 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:42:25 ; Search time 37.554 seconds
(without alignments)
1696 343 Million cell updates/sec

Title: US-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVPDHPMTDQFVTRG.....ADKKHSHGVIATGVVYLQS 474

Scoring table: BLOSUM62

Gapop 10.0, Rapexl 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_101002.*

```

1: /SID22/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/qcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2446	100.0	479	22 AAB74925	E. coli aldhyde d
2	2432	99.4	479	22 AAB74925	Novel not identifi
3	1407	57.5	476	21 AAY81486	Pseudomonas putida
4	932.5	38.1	493	23 ABP39543	Staphylococcus epi
5	931.5	38.1	479	22 AAC82331	S. epidermidis ope
6	879.5	34.0	483	22 AAO34567	Pseudomonas aerugi
7	873.5	35.7	482	22 AAU34671	E. coli cellular p
8	858.5	35.1	482	22 AAU38454	Salmonella typhi c
9	827	33.8	495	21 AAG51340	Arabidopsis thalia
10	827	33.8	528	21 AAG51349	Arabidopsis thalia

11	827	33.8	548	21 AAG51348	Arabidopsis thalia
12	826	33.8	495	21 AAG51358	Arabidopsis thalia
13	825	33.7	493	21 AAG51359	Arabidopsis thalia
14	811.5	32.8	488	23 AAR47807	Listeria monocytog
15	801.5	32.8	474	22 AAR53329	E. ribher 12-oxo la
16	797	32.6	509	22 ABB60875	Drosophila melanoq
17	796.5	32.5	442	21 AAB23350	Arabidopsis thalia
18	791.5	31.5	496	22 AAO34148	Staphylococcus aur
19	771.5	31.5	496	22 AAO37064	Staphylococcus aur
20	768.5	31.4	481	22 AAO82426	Salmonella typhi c
21	768	31.4	495	22 AAB74923	Yeast aldehyde deh
22	767.5	31.4	494	22 AAB31968	S. epidermidis ope
23	767.5	31.4	536	23 AAG33538	Staphylococcus epi
24	766.5	31.3	535	23 AAO11704	Aldehyde dehydroge
25	762.5	31.2	535	23 AAO11708	Aldehyde dehydroge
26	762.5	31.2	535	23 AAO11709	Aldehyde dehydroge
27	762	31.2	496	22 AAO34957	Enterococcus faeca
28	760.5	31.1	535	23 AAO11710	Aldehyde dehydroge
29	759.5	31.1	535	23 AAO11703	Aldehyde dehydroge
30	759.5	31.1	535	23 AAO11706	Aldehyde dehydroge
31	759.5	31.1	535	24 AAO73594	Aldehyde dehydroge
32	756.5	30.9	475	22 AAO36782	Staphylococcus aur
33	756.5	30.9	535	23 AAO11705	Aldehyde dehydroge
34	756.5	30.9	535	23 AAO11707	Aldehyde dehydroge
35	748.5	30.6	501	21 AAG36239	Arabidopsis thalia
36	748.5	30.6	501	21 AAG57413	Arabidopsis aldehy
37	737.5	30.2	488	23 AAO39401	Staphylococcus epi
38	733.5	30.0	450	22 AAO22566	Protophila melanoq
39	729	29.8	490	22 AAO36509	Pseudomonas aerugi
40	729	29.8	490	22 AAG89804	C glutamitum prote
41	729	29.8	490	22 AAG79349	Enterobacterium q
42	719.5	29.4	512	23 AAO51942	Prostate cancer as
43	719.5	29.4	512	23 AAO51942	Human aldehyde deh
44	719.5	29.4	529	22 AAO60677	Novel human diagno
45	719	29.4	497	7 AAO60456	Sequence of aldehy

ALIGNMENTS

RESULT 1
AAB74925
ID AAB74925 standard; Protein: 479 AA.

AC AAB74925;

ET 26-JUN-2001 (first entry)

DE E. coli aldhyde dehydrogenase aldA protein sequence SEQ ID NO:6.

EE Aldehyde dehydrogenase, glycerol dehydrogenase, 3-HP, glycerol,

FW foodstock, 3-hydroxypropionic acid, genetic engineering, glucose;

PW bacterial host, absorbable prosthetic device, surgical suture;

KW beta-lactam; acrylic acid; trifluoromethylated alcohol, diol,

KW polyhydroxyalkonate; copolymer; lactic acid.

XX Escherichia coli.

XX WO200116346-A1.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US23878.

XX 30 AUG 1999; 93US 0151440.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Suthers PV, Cancer DC,

XX WPI, 2001-315988/33.

XX N FSD, AAF8-083.

PT 3-hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 XX enzymes in the presence of glycerol or glucose -

PS Claim 5; Page 39-41; 63pp; English.

XX The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (I), (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhA8 gene from *Klebsiella pneumoniae* and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (I). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the *E. coli* aldehyde dehydrogenase alda,
 CC which is used in the exemplification of the present invention.

XX Sequence 479 AA;

Query Match 100.0%; Score 2446; DB 22; Length 479;
 Best Local Similarity 100.0%; Pred. No. 3.6e-217;
 Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSVPVQHPMY IDQGFVTRGDAWIDVNNIATEAVISKIPDQGAELARKALDAAEKAPQEW 60
 DB 1 MSVPVQHPMY IDQGFVTRGDAWIDVNNIATEAVISKIPDQGAELARKALDAAEKAPQEW 60
 QY 61 EALPATERASWLPRKISAGIPPEASETSALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 DB 61 EALPATERASWLPRKISAGIPPEASETSALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 QY 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPWNFPFFLIARKMAPALLTGNITVVKPSEF 180
 DB 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPWNFPFFLIARKMAPALLTGNITVVKPSEF 180
 QY 181 TTNNATAPAKIVDEIGLPRGVNVLGRGETVQGEIAGNPKVAMYSMTGSVSAGEKIMAT 240
 DB 181 TTNNATAPAKIVDEIGLPRGVNVLGRGETVQGEIAGNPKVAMYSMTGSVSAGEKIMAT 240
 QY 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 300
 DB 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 300
 QY 301 FVNRLGEAMQAVQFGNPAERNITAMGPIINAAALERVEQKVARAVEEGARVAFGKAVFG 360
 DB 301 FVNRLGEAMQAVQFGNPAERNITAMGPIINAAALERVEQKVARAVEEGARVAFGKAVFG 360
 QY 361 KGYYPPTLLLDVRQEMSIMHEETFGPVLPPVAFADILEDAISMANDSYGLTSSITYQNL 420
 DB 361 KGYYPPTLLLDVRQEMSIMHEETFGPVLPPVAFADILEDAISMANDSYGLTSSITYQNL 420
 QY 421 NVAMKALKGLKFGETVINRENFEAMQGFHAGWKKSGIGSADGKHGLHGLYLOTQVYYLQS 479
 DB 421 NVAMKALKGLKFGETVINRENFEAMQGFHAGWKKSGIGSADGKHGLHGLYLOTQVYYLQS 479

RESULT 2

AAU29333

ID AAU29333 standard; Protein; 479 AA.

XX

AC AAU29333;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel mar regulated protein (NIMR) #5.
 XX mar regulated polypeptide; NIMR; microbial infection; antibacterial.
 XX *Escherichia coli*.
 XX W-200179776-A2.
 XX 27-SEP-2001.

XX 38-MAR-2001; 2001WO-0507478.
 XX 10-MAR-2000; 2000QS-148362P.
 XX (TUFT) TUFTS COLLEGE.

XX Levy SB, Barbosa TM, Alekshun MN;
 XX WPI: 2001-602769/68.
 XX N-PSDB; AAS46235.

XX Identifying compounds that modulate a newly identified mar regulated
 PT polypeptide activity, useful as antimicrobial compounds, involves
 PT contacting the polypeptide with a test compound

XX Disclosure; Page 160-161; 526pp; English.

XX The invention relates to a method of identifying compounds that modulate
 CC a newly identified mar regulated (NIMR) polypeptide activity. The method
 CC comprises contacting an NIMR polypeptide with a test compound under
 CC interaction conditions, determining the ability of the compound to
 CC modulate the activity or expression of the polypeptide, and selecting the
 CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
 CC of microbial infections, and in screening for modulators of NIMR
 CC expression and activity. These modulators can be used to reduce the
 CC infectivity of a microbe on a surface, and the virulence of a microbe in
 CC a subject suffering from an infection. AAU29333-AAU29479 represent
 CC *Escherichia coli* NIMR amino acid sequences of the invention.

XX Sequence 479 AA;

Query Match 99.4%; Score 2442; DB 22; Length 479;
 Best Local Similarity 99.6%; Pred. No. 3.2e-216;
 Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVPVQHPMY IDQGFVTRGDAWIDVNNIATEAVISKIPDQGAELARKALDAAEKAPQEW 60
 DB 1 MSVPVQHPMY IDQGFVTRGDAWIDVNNIATEAVISKIPDQGAELARKALDAAEKAPQEW 60
 QY 61 EALPATERASWLPRKISAGIPPEASETSALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 DB 61 EALPATERASWLPRKISAGIPPEASETSALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 QY 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPWNFPFFLIARKMAPALLTGNITVVKPSEF 180
 DB 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPWNFPFFLIARKMAPALLTGNITVVKPSEF 180
 QY 181 TTNNATAPAKIVDEIGLPRGVNVLGRGETVQGEIAGNPKVAMYSMTGSVSAGEKIMAT 240
 DB 181 TTNNATAPAKIVDEIGLPRGVNVLGRGETVQGEIAGNPKVAMYSMTGSVSAGEKIMAT 240
 QY 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 400
 DB 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 400
 QY 301 FVNRLGEAMQAVQFGNPAERNITAMGPIINAAALERVEQKVARAVEEGARVAFGKAVFG 460
 DB 301 FVNRLGEAMQAVQFGNPAERNITAMGPIINAAALERVEQKVARAVEEGARVAFGKAVFG 460
 QY 361 KGYYPPTLLLDVRQEMSIMHEETFGPVLPPVAFADILEDAISMANDSYGLTSSITYQNL 420
 DB 361 KGYYPPTLLLDVRQEMSIMHEETFGPVLPPVAFADILEDAISMANDSYGLTSSITYQNL 420

QY 421 NVAMKAIKGLKPESEIYINKEFEAMQGFHAGWRKSGDLSALQKHHL45YLTQGVVYIUS 479
 Dd 421 NVAMKAIKGLKPESEIYINKEFEAMQGFHAGWRKSGDLSALQKHHL45YLTQGVVYIUS 479

RESULT 3
 AAY81486
 ID AAY81486 standard; Protein; 476 AA
 XX
 AC AAY81486;
 XX
 XX 03-JUL-2000 (first entry)
 DT
 DE
 XX Pseudomonas putida aldehyde dehydrogenase.
 XX
 XX Aldehyde dehydrogenase, NAD⁺ dependent; oxidation, 2-oxalaldehyde;
 KW hydroxypyruvaldehyde; glyceraldehyde; methylglyoxal; hydroxypyruvic acid.
 KW
 XX Pseudomonas putida.
 OS
 XX TP200006053-A
 PN
 XX 29-FEB-2000.
 PD
 XX
 XX 24-AUG-1998; 98JP-0236691.
 PF
 XX
 XX 24-AUG-1998; 98JP-0236691.
 PR
 XX
 XX (CHIK-) ZH CHIKYU KANKYO SANGYO GJITSU KENKYU.
 PA
 XX
 XX WPI; 2000-249678/22.
 DR
 DR N-PSDB; AAA07084.
 DR
 XX New aldehyde dehydrogenase gene - for enzymatic synthesis of serine or
 PT cysteine from glycerol
 PT
 XX
 XX Claim 2; Page 8-9; 11pp; Japanese
 PS
 XX
 XX This sequence represents a novel aldehyde dehydrogenase from
 CC Pseudomonas putida. This catalyses reactions such as the NAD⁺-dependent
 CC oxidation of hydroxypyruvaldehyde to hydroxypyruvic acid and reacts
 CC specifically with 2-oxalaldehydes such as hydroxypyruvaldehyde,
 CC glyceraldehyde and methylglyoxal. The protein was initially purified
 CC from a culture of P. putida and subjected to tryptic cleavage. The
 CC fragments thus obtained were sequenced, and the sequence information used
 CC to design primers AAA07084-A07086. These primers were used to amplify a
 CC aldehyde dehydrogenase PCR product. The fragment was used in the
 CC detection of a full-length DNA encoding aldehyde dehydrogenase in a
 CC genomic library. This was amplified using primers AAA07087-A07088, and
 CC inserted into a vector for transformation of cells. Percombinantly
 CC produced Pseudomonas putida aldehyde dehydrogenase may be used for the
 CC synthesis of serine or cysteine from glycerol.
 XX
 SQ Sequence 476 AA;

Query Match 57.5%; Score 1407; DB 21; Length 476;
 Best Local Similarity 58.9%; Pred. No. 3.9e-12;
 Matches 277; Conservative 77; Mismatches 112; Indels 4; Gaps 3,
 10 YINGQFVTVWQFJAWIIVNFALEAVISHPHESJAFKAKLIDAAFAAPQWRAIPATEPA 69
 Dd 9 YIDNAFVSEG--LLEVYNPANALLGRVPSPQVERALAAKAKQAKWAAPALERA 66
 QY 70 SWLRKISAGIRKASLEISALIVEEGGKTLQLAKEVEVAFTATYIHYMAEWAPPYEGEITQS 124
 Dd 67 GYLQIAAKAVPANADRIARIITQEGGKVPGLAQGVNFTADYLDYMAEWARRLEGEVLTS 126
 QY 130 DRPGENILFKKALGVTTGILWNFPFELIARKKAPALLTGNTIVIKPSEFTNNATAFA 189
 Dd 127 DRANGHTFLMRKPLVWAGILWNFPFELIARKKAPALLTGNTIVIKPSEFTNINYEFA 186
 QY 190 KIVDRIGIPRVFNIVIGFQTVQELAGNFKVAMVSMQSVSAKCKIMATAKNIKVC 249

Dd 187 KIVAFITLPA3GVNVNVRCKGASVSH4LSSHAGLIIHVSFTGSVAISAPIMAAAAFPIIKLIN 246
 QY 256 LELGSKAIVIMUQANILFLAVKALIVISFVINSGLQWVNWAPVYVQKTHIQFVNFHIGRAM 309
 Dd 247 LELGSKAPAIYLAADADLALATKATIVASPVNTGQVWNCAPVYVVARVADAFYDKVAQAM 306
 QY 310 QAVQFGNFAENQJAMGHLINAAALEPVEKVAFAVFEAFVAFGRKAVF--GKSYVYPPPI 368
 Dd 307 AATPYGHSPPAGLIMSLINLQKVAJWVPTAVGGRAGIVTGQVAVAGLQACFHYQPT 366
 QY 369 LLDVVRQMSIMHEETFGVLPVAVFDLEDAISMANDSYGLTSSITVQNLNVAMKAIK 428
 Dd 367 VLA-CAADMEIMRKFIQGVLP IQWVVDLEALALANDSEYGLTSSITVQNLNVAMKAIK 425
 QY 429 GLKFGETVINRENFEAMOGFHAGWRKSGIGGADCKHGLHGYLTQVVYLO 478
 Dd 426 EIDFGETVINRENFEAMOGFHAGWRKSGIGGADCKHGLYETPTQVVYLO 475

RESULT 4
 ABP39543
 ID ABP39543 standard; Protein; 493 AA.
 XX
 AC ABP39543;
 XX
 XX 24-JUL-2002 (first entry)
 DT
 DE
 XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4388.
 XX
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial, gene therapy.
 KW
 XX Staphylococcus epidermidis.
 OS
 XX US6380370-B1.
 PN
 XX 30-APP-2002.
 PD
 XX
 XX 13-AUG-1998; 98US-0134001.
 PF
 XX
 XX 14-AUG-1997; 97US-055779P.
 PR
 XX 08 NOV-1997. 97US-054964P.
 XX
 XX (GENE-) GENOME HPPAPPHIUS Corp.
 XX
 XX Doucette-Stamm LA, Bush D;
 XX
 XX WPI; 2002-381255/41.
 DR
 DR N-PSNR; ARN92088.
 DR
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 ET polypeptide, useful for diagnosing and treating bacterial infections -
 FI
 XX Disclosure; SEQ ID 4388; 267pp; English.
 PS
 XX
 XX ARN9538 to ARN9374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ARN9124 to ARN9796. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX
 SQ Sequence 493 AA;

Query Match 38.1%; Score 932.5; DB 23; Length 493;
 Best Local Similarity 38.5%; Pred. No. 2.9e-77;
 Matches 161; Conservative 108; Mismatches 186; Indels 1; Gaps 1;

QY 9 MYIDGQVTVRGDAWIDVNVNPAEAVLSRIPIESJALAKAPKAIHAAEAPPEWALPALER 68
 DB 24 LFINNEFTESQSKETMDVINPATGEAFDTITLATEEEVNDALTEKSOQALEWERVPQPTR 83
 QY 69 ASWLKRTSAGIRERASKISALIVBEGSKIQQLAEVEVAFADYIDYMAEWARRYEGEIIQ 128
 DB 84 AEHVKLLIPLLEKNRDELAOLYVKEOGKTLAQAYGETDKSTISFDYMTSLSMSKGVVLQ 143
 QY 129 SDRPGENILFLKRALGVTTGILPWNPFFFLIARKMAPALLTNTIVIKPSEFITTNAIAF 188
 DB 144 NSIANETIQINKPIGVGTAGVFNWNPAPILVLMKKVIPAIVTQGVVWIKPSEETILLTKL 203
 QY 189 AKIVDEICLPRGVNVLICRGETVGOELAGNPKVAMVMTGSVSAAGEKIMATAAKNITKV 248
 DB 204 AELFRASITPAGLFQIVPVTGETVGTQLASHKULQLSLGSMKPAKSVYENAAVTVKV 263
 QY 249 CLEGGKAPATVMDADLELAVKAIIVDSRVINSQVNCNAERVYKGIYDQFVNRUGEA 308
 DB 264 NLEGGNAPVIVTSNADLUKAVNYIVTAKINAGQVCTCPKIFVHEDVHDLNKNVTSK 323
 QY 309 MOAVQFCNPAERNIAMGPIINAALEPVEQKVAPEVCAVAFGKAVEGKGVYYPPT 368
 DB 324 MKSLTVGDPDEN-TDYGATINQKQDSIHEKVDIAIKNGATILMTGGHOLKHHGFFYAPT 382
 QY 369 LLLDVROEMSIMHEETFGVPLPVYVAFDTLEDAISMANDSDYGITSSIVTONTLNAMKAIR 428
 DB 383 VLDNVRKDYVVKPFDIFGCPVLAITTYPDFFQVIEDANDINAGLSSYIFSENITEVMTATE 442
 QY 429 GLKPGFTYINPFFFAWQPHACWPKSGTGATGKHGIHSLYLTQVAVYLQ 478
 DB 443 RLKPGFYVYANCAEEVYNGYHAGWRESGLGADGTHGFEFYNTTVSVIR 492

RESULT 5

AA082931
 ID AA082931 standard; Protein: 479 AA.
 XX
 AC
 XX
 AA082931;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:2956.
 XX
 KW Staphylococcus epidermidis SK1 strain; infection; diagnosis;
 XX
 KW vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN W0200134809-A2.
 XX
 PD 17 MAY 2001
 XX
 PF 09-NOV-2000; 2000W0-US-0787
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimberly WJ;
 XX
 DR WPI; 2001-316495/33.
 XX
 DR N-PSDR; AAH53781.
 XX

Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 PT useful for vaccination against infections, e.g. endocarditis.

PS Claim 18; Page 773; 2188pp. English.

XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAH51454 to AAH53120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH54971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 479 AA;

Query Match 38.1%, score 941.5, Db 22; Length 479;
 Best Local Similarity 38.5%; Pred. No. 356 77;
 Matches 181; Conservative 109; Mismatches 179; Indels 1; Gaps 1;
 QY 9 MYIDGQVTVRGDAWIDVNVNPAEAVLSRIPIESJALAKAPKAIHAAEAPPEWALPALER 68
 DB 10 LFINNEFTESQSKETMDVINPATGEAFDTITLATEEEVNDALTEKSOQALEWERVPQPTR 69
 QY 69 ASWLKRTSAGIRERASKISALIVBEGSKIQQLAEVEVAFADYIDYMAEWARRYEGEIIQ 128
 DB 70 AEHVKLLIPLLEKNRDELAOLYVKEOGKTLAQAYGETDKSTISFDYMTSLSMSKGVVLQ 129
 QY 129 SDRPGENILFLKRALGVTTGILPWNPFFFLIARKMAPALLTNTIVIKPSEFITTNAIAF 188
 DB 130 NSIANETIQINKPIGVGTAGVFNWNPAPILVLMKKVIPAIVTQGVVWIKPSEETILLTKL 189
 QY 189 AKIVDEICLPRGVNVLICRGETVGOELAGNPKVAMVMTGSVSAAGEKIMATAAKNITKV 248
 DB 190 AELFRASITPAGLFQIVPVTGETVGTQLASHKULQLSLGSMKPAKSVYENAAVTVKV 249
 QY 249 CLEGGKAPATVMDADLELAVKAIIVDSRVINSQVNCNAERVYKGIYDQFVNRUGEA 308
 DB 250 NLEGGNAPVIVTSNADLUKAVNYIVTAKINAGQVCTCPKIFVHEDVHDLNKNVTSK 309
 QY 309 MOAVQFCNPAERNIAMGPIINAALEPVEQKVAPEVCAVAFGKAVEGKGVYYPPT 368
 DB 310 MKSLTVGDPDEN-TDYGATINQKQDSIHEKVDIAIKNGATILMTGGHOLKHHGFFYAPT 368
 QY 369 LLLDVROEMSIMHEETFGVPLPVYVAFDTLEDAISMANDSDYGITSSIVTONTLNAMKAIR 428
 DB 369 VLDNVRKDYVVKPFDIFGCPVLAITTYPDFFQVIEDANDINAGLSSYIFSENITEVMTATE 428
 QY 429 GLKPGFTYINPFFFAWQPHACWPKSGTGATGKHGIHSLYLTQVAVYLQ 478
 DB 429 PLKPGFYVYANCAEEVYNGYHAGWRESGLGADGTHGFEFYNTTVSVIR 478

RESULT 6

AA033562
 ID AA033562 standard; Protein: 483 AA
 XX
 AC
 XX
 AA033562;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa cellular proliferation protein #6.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.

XX W:200170955-A2.

PD 27-SEP-2001.

Query Match 35.78; Score 873.5; DB 22; Length 482;
 Best Local Similarity 39.58; Pred. No. 8e-72;
 Matches 182; Conservative 91; Mismatches 187; Indels 1; Gaps 1.

QY 11 IDGGFVTRGDAWIDVYNPATIAVLSKIPGQAEDAKKALDAAEKAPQEWALPALEAS 70
 Db 15 INGEWLDANNGEADVTNPANGDKIGSVPMKGADETPAALDANRALPAWRLTAKERAT 74
 QY 71 WLKISACTIRRAEISALIVERGKIQQLAEVEAFTAIYIDYMAFWAPRYEGFTQSD 130
 Db 75 ILRWFLNLMHEDDILARLMTLEGGKPLAEAKGEISYAASFIEWEAECKRIGYDTIFCH 134
 QY 131 RPSGNILFLKRALGVTGILPWPMPFELIARKMAPALLTNTIVIKPSEFTNNALATAK 190
 Db 135 QADRLIVIKQPIGVTAATTPWNPFAAMITRKAGPALAAGCTMWLKPASQTPFSALALAE 194
 QY 191 IVDFIGLPRGVFNILVGRGETVQGLAGNPKVAMVMTGVSAGSEKIMATAAKNITKVCL 250
 Db 195 LAIRAGVFAGVFNVTGSAGAVGNELTNPVLRKLSFTGSTELGRMLMECCAKDKIKVSL 254
 QY 251 ELGGKAPALVMDADLELAWKAIVDSRVINSQVCMCAEVYVQGYIDQFVNRIGEAHQ 310
 Db 255 ELGGMNAPFIVFDGADLEKAVESALASKFRNAGQTCVANKLYVQGVYDURFAEKLQQA 314
 QY 311 AVQFQFAEKNDIAMGFLJNAAALERVEOKVARAVEEGARVAFGKVGKGYVYPPTIL 370
 Db 315 KLHIGDGLD NGVTIGLIDEKAVAKVEEHIADALEKGPVVCGKAHERGGNFQPTIL 373
 QY 371 LDVQKQMSIMHEETFGVPLDVVAFDTLENAISMANSDYGLTSSYTONLNVAMKAIKOL 430
 Db 374 VDPANAKVSKEETFGPLAPLFRKDEADVIAQANDTEFLAAYFYARDLSVFRVGEAL 433
 QY 431 KFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLO 471
 Db 434 EYGVINGTGIISNEVAPFGIKASGLRGSKYGIETYLE 474

RESULT 8
 AAU38454
 ID AAU38454 standard; protein: 482 AA.

AAU38454;

14-FEB-2002 (first entry)

Salmonella typhi cellular proliferation protein #345.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Salmonella typhi.

WO200170955-A2.

27-SEP-2001.

21 MAR 2001; 2001WO-US094180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-20727P.

23-OCT-2000; 2000US-24578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269408P.

(ELIT-) ELITRA PHARM INC.

Haselebeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu RH;

WPI; 2001-611495/70.

N-PSDH; AAS56313.

New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids

Example 3; Seq ID No 14047; 51pp; English.

The invention relates to antisense inhibitors of genes essential to
 prokaryotic cellular proliferation, their use in identifying the
 genes, their use in the discovery of novel antibiotics, the essential
 genes themselves and the encoded proteins, the prokaryotes used are
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
 invention is also useful for the identification of potential new targets
 for antibiotic development. The antisense nucleic acids can also be used
 to identify proteins used in proliferation, to express these proteins,
 and to obtain antibodies capable of binding to the expressed proteins.
 The proteins can be used to screen compounds in rational drug discovery
 programmes. The antisense nucleic acid sequence is also useful to screen
 for homologous nucleic acids which are required for cell proliferation in
 a wide variety of organisms. The present sequence represents an
 essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part
 of the printed specification, but was obtained in electronic
 format directly from WIPO at

ftp.wipo.int/pub/published_jpt_sequences.

Sequence 482 AA;

Query Match 35.18; Score 858.5; DB 22; Length 482;
 Best Local Similarity 39.28; Pred. No. 2e-70;
 Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

QY 10 YIDGQFVTRGDAWIDVYNPATIAVLSKIPGQAEDAKKALDAAEKAPQEWALPALEAS 69
 Db 14 FIDGWRDARGGVIVPSNPANKPLGNVFKMCAETRTDIAINAKRALPAWRLTAKERAT 74
 QY 70 SWLRKISAGIRERASEISALIVERGKIQQLAEVEAFTAIYIDYMAFWAPRYEGFTQSD 129
 Db 74 NLRWFLNLMHEDDILARLMTLEGGKPLAEAKGEISYAASFIEWEAECKRIGYDTIFCH 134
 QY 130 DRPCENILFLKRALGVTGILPWPMPFELIARKMAPALLTNTIVIKPSEFTNNALATAK 189
 Db 134 BQDCKRLIVIKQPIGVTAATTPWNPFAAMITRKAGPALAAGCTMWLKPASQTPFSALALAE 194
 QY 190 KIVDFIGLPRGVFNILVGRGETVQGLAGNPKVAMVMTGVSAGSEKIMATAAKNITKVCL 249
 Db 194 ELAQRACILLAGVFNVTGSAGAVGNELTNPVLRKLSFTGSTELGRMLMECCAKDKIKVSL 254
 QY 250 LEUGKAPALVMDADLELAWKAIVDSRVINSQVCMCAEVYVQGYIDQFVNRIGEAHQ 309
 Db 254 LEUGGNAPFIVFDGADLEKAVESALASKFRNAGQTCVANKLYVQGVYDURFAEKLNQAV 314
 QY 310 QAVQFQFAEKNDIAMGFLJNAAALERVEOKVARAVEEGARVAFGKVGKGYVYPPTIL 369
 Db 314 NKLAVGDLGDA IVALGPIIDKAVAKVOEHLAELKCARVLDGEAKRLGGNFQPTIL 472
 QY 370 LLDVQKQMSIMHEETFGVPLDVVAFDTLENAISMANSDYGLTSSYTONLNVAMKAIKOL 429
 Db 373 LAIVTNRKAVAKIETFGPLAPLFRKDEADVIAQANDTEFLAAYFYARDLSVFRVGEAL 432
 QY 430 KFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLO 471
 Db 433 EYGVINGTGIISNEVAPFGIKASGLRGSKYGIETYLE 474

RESULT 9

AAG51350

ID AAG51350 standard; protein: 495 AA.

AC AAG51350;

XX 18-OCT-2000 (first entry)

us-09-830-751-6.rag

Tue Jun 24 10:42:22 2003

XX	Arabidopsis thaliana protein fragment SEQ ID NO: 65163.	PR	21-JUN-1999;	99US-0139817.
DE		PR	22-JUN-1999;	99US-0139899.
XX		PR	23-JUN-1999;	99US-0140353.
XX		PR	23-JUN-1999;	99US-0140354.
KW	Protein identification: signal transduction pathway; metabolic pathway;	PR	24-JUN-1999;	99US-0140695.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	28-JUN-1999;	99US-0140823.
KW	termination sequence.	PR	29-JUN-1999;	99US-0140991.
XX		PR	30-JUN-1999;	99US-0141287.
OS	Arabidopsis thaliana.	PR	01-JUL-1999;	99US-0141842.
XX		PR	02-JUL-1999;	99US-0142055.
PN	EP1033405-A2.	PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PD	06-SEP-2000.	PR	09-JUL-1999;	99US-0142920.
XX		PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143342.
XX		PR	14-JUL-1999;	99US-0143624.
XX		PR	15-JUL-1999;	99US-0144005.
XX		PR	16-JUL-1999;	99US-0144085.
XX		PR	16-JUL-1999;	99US-0144086.
XX		PR	19-JUL-1999;	99US-0144325.
XX		PR	19-JUL-1999;	99US-0144331.
XX		PR	19-JUL-1999;	99US-0144332.
XX		PR	19-JUL-1999;	99US-0144333.
XX		PR	19-JUL-1999;	99US-0144335.
XX		PR	20-JUL-1999;	99US-0144352.
XX		PR	20-JUL-1999;	99US-0144632.
XX		PR	20-JUL-1999;	99US-0144884.
XX		PR	21-JUL-1999;	99US-0144814.
XX		PR	21-JUL-1999;	99US-0145086.
XX		PR	21-JUL-1999;	99US-0145088.
XX		PR	22-JUL-1999;	99US-0145085.
XX		PR	22-JUL-1999;	99US-0145087.
XX		PR	22-JUL-1999;	99US-0145089.
XX		PR	22-JUL-1999;	99US-0145192.
XX		PR	23-JUL-1999;	99US-0145145.
XX		PR	23-JUL-1999;	99US-0145218.
XX		PR	23-JUL-1999;	99US-0145224.
XX		PR	23-JUL-1999;	99US-0145276.
XX		PR	26-JUL-1999;	99US-0145913.
XX		PR	27-JUL-1999;	99US-0145918.
XX		PR	27-JUL-1999;	99US-0145919.
XX		PR	27-JUL-1999;	99US-0145951.
XX		PR	28-JUL-1999;	99US-0146386.
XX		PR	02-AUG-1999;	99US-0146388.
XX		PR	02-AUG-1999;	99US-0146389.
XX		PR	02-AUG-1999;	99US-0147038.
XX		PR	03-AUG-1999;	99US-0147204.
XX		PR	04-AUG-1999;	99US-0147302.
XX		PR	05-AUG-1999;	99US-0147360.
XX		PR	05-AUG-1999;	99US-0147361.
XX		PR	06-AUG-1999;	99US-0147303.
XX		PR	06-AUG-1999;	99US-0147416.
XX		PR	09-AUG-1999;	99US-0147493.
XX		PR	09-AUG-1999;	99US-0147935.
XX		PR	09-AUG-1999;	99US-0148171.
XX		PR	10-AUG-1999;	99US-0148319.
XX		PR	11-AUG-1999;	99US-0148341.
XX		PR	12-AUG-1999;	99US-0148565.
XX		PR	13-AUG-1999;	99US-0148684.
XX		PR	13-AUG-1999;	99US-0149368.
XX		PR	16-AUG-1999;	99US-0149175.
XX		PR	17-AUG-1999;	99US-0149426.
XX		PR	18-AUG-1999;	99US-0149722.
XX		PR	20-AUG-1999;	99US-0149723.
XX		PR	20-AUG-1999;	99US-0149929.
XX		PR	23-AUG-1999;	99US-0149902.
XX		PR	23-AUG-1999;	99US-0149930.
XX		PR	25-AUG-1999;	99US-0150566.
XX		PR	26-AUG-1999;	99US-0150884.
XX		PR	27-AUG-1999;	99US-0151065.
XX		PR	27-AUG-1999;	99US-0151066.

```

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161459.
PR 26-OCT-1999; 99US-0161460.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match. 33.8%; Score 827; DB 21; Length 495;
Best local Similarity 37.8%; Pred. No. 1.7e-67;
Matches 185; Conservative 90; Mismatches 193; Indels 22; Gaps 3;

QY 9 MYIDGQVFWR-----GDWID-----VNPATEAVISRIPDQAEAR 47
DB 1 MSMDAQSSEKLSGLKIGLGGKWDSDYDKIKNVNPATGEIADVACMGKFTN 60
QY 48 KAIIDAFRAQDFEALPATERASWLRKISAGIPEPASEISALIVEEGKIOOLAEEVAF 107
DB 61 DALASSYFAITWSRPTAGFSPKVPFPWYDIIAHKEELGSLITLQCKPLKEALGEVAY 120
QY 108 TADYIDYMAFWARHYSFIIQSNPCENILLPKRALGVTTGLLPWNFFPFLIARKMAPAL 167
DB 121 GASFIYYAEAKRVYGDIIIPNLSDRLVLKQPVGVGALITWNFPLAMITRKVGFPAL 180
QY 168 LTGNTIVIKPSFPTNNATAFAKIVDEIGLPGVFNVLGRGETVGGFAGNPKVAMVSM 227
DB 181 ASQCTVVVKSPLTPIALAAAEALQAGVPIGALNVVGNAPFGLDALLSPQVRKITP 240
QY 228 TGSVSAGKIMATAAKNIITKVCELGSGKAPALVMDADALELAKVALIVSRVINSQVCNC 287
DB 241 TGSTAVGKKLMAAAPTVKKSLGNGNAPSIVFDADLDVAVKGTILAAKFPNSGQTCVC 300

```

```

QY 288 AERVVYVKGILYQFVNRLGEAMQAVQFQNFPAERNNHIAHPLINAAALERVQKAVAEV 447
DB 301 ANPVLVQNGIYIKFAFAFSEAVUKLFVJHGFECITGTPLINAAVQKVFTEVQAVSR 459
QY 348 GARVAFGGRKAVGKGYYYPTLLAVROHMSIMHEETEDVGFVAVPDIIEAISMANNIS 407
DB 360 GAKTITGGEPHSIGMTFFVPTVIVAVGNNMINSKFFIETPVADITPPTTEPATRIANIT 419
QY 408 DGLTSSLYIQNLINAMAKALEKFEQYIINKENFAMQGEHAIWKKSLGGLAALHGLH 467
DB 420 LACLAAYIPTNSVQSWRPVITALEYGLVAVNIEELITTEVAVATGAVKATLDEKPKYMG 479
QY 468 GYLQTOVVYL 477
DB 480 EYLEIKYVCL 489

RESULT 10
AAC51349
IP AAC51349 standard; Protein; 528 AA.
XX AC
XX AAG51349;
XX 18 OCT 2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 65162.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 05-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.

```

PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137222.
PR 04-JUN-1999; 990S-0137222.
PR 07-JUN-1999; 990S-0137222.
PR 08-JUN-1999; 990S-0137222.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139453.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139829.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 06-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142820.
PR 12-JUL-1999; 990S-0142877.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143544.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 23-JUL-1999; 990S-0145132.
PR 23-JUL-1999; 990S-0145135.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147004.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147402.
PR 05-AUG-1999; 990S-0147432.
PR 05-AUG-1999; 990S-0147460.
PR 06-AUG-1999; 990S-0147463.
PR 06-AUG-1999; 990S-0147416.
PR 06-AUG-1999; 990S-0147493.
PR 06-AUG-1999; 990S-0147495.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149922.
PR 23-AUG-1999; 990S-0149940.
PR 25-AUG-1999; 990S-0150565.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151403.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151830.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155135.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158469.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 14-OCT-1999; 990S-0159684.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160880.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 25-OCT-1999; 990S-0161584.
PR 25-OCT-1999; 990S-0161585.
PR 25-OCT-1999; 990S-0161586.
PR 25-OCT-1999; 990S-0161587.
PR 25-OCT-1999; 990S-0161588.
PR 25-OCT-1999; 990S-0161589.
PR 25-OCT-1999; 990S-0161590.
PR 25-OCT-1999; 990S-0161591.
PR 25-OCT-1999; 990S-0161592.
PR 25-OCT-1999; 990S-0161593.

[illegible]

```

PR 19-JUL-1999; 99US-0144333
PR 19-JUL-1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335
PR 20-JUL-1999; 99US-0144352
PR 20-JUL-1999; 99US-0144632
PR 20-JUL-1999; 99US-0144884
PR 21-JUL-1999; 99US-0144814
PR 21-JUL-1999; 99US-0145086
PR 21-JUL-1999; 99US-0145088
PR 21-JUL-1999; 99US-0145085
PR 22-JUL-1999; 99US-0145087
PR 22-JUL-1999; 99US-0145089
PR 22-JUL-1999; 99US-0145192
PR 23-JUL-1999; 99US-0145145
PR 23-JUL-1999; 99US-0145218
PR 23-JUL-1999; 99US-0145224
PR 26-JUL-1999; 99US-0145276
PR 27-JUL-1999; 99US-0145913
PR 27-JUL-1999; 99US-0145918
PR 27-JUL-1999; 99US-0145919
PR 28-JUL-1999; 99US-0145951
PR 02-AUG-1999; 99US-0145984
PR 02-AUG-1999; 99US-0145989
PR 02-AUG-1999; 99US-0146389
PR 03-AUG-1999; 99US-0147038
PR 04-AUG-1999; 99US-0147204
PR 04-AUG-1999; 99US-0147302
PR 05-AUG-1999; 99US-0147192
PR 05-AUG-1999; 99US-0147260
PR 06-AUG-1999; 99US-0147403
PR 06-AUG-1999; 99US-0147416
PR 09-AUG-1999; 99US-0147493
PR 09-AUG-1999; 99US-0147935
PR 10-AUG-1999; 99US-0148171
PR 11-AUG-1999; 99US-0148319
PR 12-AUG-1999; 99US-0148341
PR 13-AUG-1999; 99US-0148565
PR 13-AUG-1999; 99US-0148684
PR 16-AUG-1999; 99US-0149368
PR 17-AUG-1999; 99US-0149175
PR 18-AUG-1999; 99US-0149426
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149723
PR 23-AUG-1999; 99US-0149902
PR 23-AUG-1999; 99US-0149930
PR 25-AUG-1999; 99US-0150566
PR 26-AUG-1999; 99US-0150884
PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151062
PR 27-AUG-1999; 99US-0151080
PR 30-AUG-1999; 99US-0151303
PR 31-AUG-1999; 99US-0151438
PR 01-SEP-1999; 99US-0151940
PR 07-SEP-1999; 99US-0152363
PR 10-SEP-1999; 99US-0153070
PR 13-SEP-1999; 99US-0153758
PR 13-SEP-1999; 99US-0154018
PR 16-SEP-1999; 99US-0154039
PR 20-SEP-1999; 99US-0154779
PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157753
PR 06-OCT-1999; 99US-0157865
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232
PR 12-OCT-1999; 99US-0158369
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159294

```

```

PR 13-OCT-1999; 99US-0159295
PR 14-OCT-1999; 99US-0159329
PR 14-OCT-1999; 99US-0159330
PR 14-OCT-1999; 99US-0159331
PR 14-OCT-1999; 99US-0159637
PR 14-OCT-1999; 99US-0159638
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741
PR 21-OCT-1999; 99US-0160767
PR 21-OCT-1999; 99US-0160768
PR 21-OCT-1999; 99US-0160770
PR 21-OCT-1999; 99US-0160814
PR 21-OCT-1999; 99US-0160815
PR 22-OCT-1999; 99US-0160980
PR 22-OCT-1999; 99US-0160981
PR 22-OCT-1999; 99US-0160981
PR 25-OCT-1999; 99US-0161404
PR 25-OCT-1999; 99US-0161405
PR 25-OCT-1999; 99US-0161406
PR 26-OCT-1999; 99US-0161359
PR 26-OCT-1999; 99US-0161360
PR 26-OCT-1999; 99US-0161361
PR 26-OCT-1999; 99US-0161920
PR 28-OCT-1999; 99US-0161992
PR 28-OCT-1999; 99US-0161993
PR 29 OCT 1999; 99US 0162142

Query Match 33.8%; Score 827; DB 21; Length 548;
Best Local Similarity 37.8%; Pred. No. 1.9e-67;
Matches 185; Conservative 90; Mismatches 194; Indels 22; Gaps 3;

QY 9 MYIDQFVWTP-----GDWID-----VVPATEAVISIPGQADAP 47
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Qb 54 MSMAQSVSEKLPSSGLLPGLDQKWLISYFNKTIKVNPAIGIADVACMGKTN 113
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 48 KATDAEPAPQPFWEALPAIEPASWIRKISAGTTPPASHISALIVRSGKIQQLAFVPAF 107
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 DATASSYEAFTSWRPLTAGESKVLPPWYDILLIAHKEFLGQLITLPGSKPIKEATGEVAY 173
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 TADYIDYMAEWAPPEYPTIGSRPRNNTIIFKALGVTTGTLFWNPFFIAPKNAPAL 167
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Lb 174 GASFLEYALLAKPVYGLIIFNLDEPLVIRKGVSVCAITWNEFLAMITKRVGFAL 233
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 LGNTIVIKISLFTNNALAPAKIVLEIGLIGOVENIVDGRKLVGLELAKNFKVAMVSM 227
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Eb 234 ASPTVWVVFESPLTFITLAAAFALGAGVTPGALNVMMENADPTIDPATITSPQVPKTF 263
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 238 LCSVSAGCKIMAJAAKNIKVLELGSKAFAVMWAWGLFAYPAIVFSAVHSQAVQNC 287
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Eb 294 FGSTAVGPKLMAAAAFVFEVSTFISENAVSIVTFGAGVAVVTTTAAFTASQQLQVC 353
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 APPVYVAKGTIDQFVNPIQTEAMQAVQGNAPFNDIANGPTINAAALPPVQKVPAAVEE 347
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Eb 354 ANPLVLDGIVYKFAEAFSEAVQKIFVQD-GFPSTLGGPIINIAAVKVEIEVQAVSK 412
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 GAPVAFGKAVRSGKYYPPTLLLDVROEMSMHEETPGVPLVPAVEDTLBDALSMANUS 407
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 413 GAKIIGCKPHSTGMTFYPTVTPVSDNMIMSKKEIEFGVPAPLTPKTEKTDPAIANT 472
  | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 408 DYGLISSIYTONLNVAMKAIKGLKPGETYINKENFEAMQGHGACPKSGIGGAGKHGLH 467
  | | | | | | | | | | | | | | | | | | | | | | | | | |
Eb 473 TAGLAAYIFTNVSQESWVFEALELGLVGVNPFELISTFVAFPGVAFSGTGPFGSGSGMD 542
  | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
AAC23358
ID AAG43558 standard; Protein; 495 AA.
XX
AC AAC23358;

```

XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 26911.
DE
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
XX 05-MAR-1999; 990S-0123180.
XX 09-MAR-1999; 990S-0123548.
XX 23-MAR-1999; 990S-0125788.
XX 25-MAR-1999; 990S-0126264.
XX 29-MAR-1999; 990S-0126785.
XX 01-APR-1999; 990S-0127462.
XX 06-APR-1999; 990S-0128234.
XX 08-APR-1999; 990S-0128714.
XX 16-APR-1999; 990S-0129845.
XX 19-APR-1999; 990S-0130077.
XX 21-APR-1999; 990S-0130449.
XX 23-APR-1999; 990S-0130510.
XX 23-APR-1999; 990S-0130891.
XX 28-APR-1999; 990S-0131449.
XX 30-APR-1999; 990S-0132048.
XX 30-APR-1999; 990S-0132407.
XX 04-MAY-1999; 990S-0132484.
XX 05-MAY-1999; 990S-0132485.
XX 06-MAY-1999; 990S-0132486.
XX 06-MAY-1999; 990S-0132487.
XX 07-MAY-1999; 990S-0132863.
XX 11-MAY-1999; 990S-0134256.
XX 14-MAY-1999; 990S-0134218.
XX 14-MAY-1999; 990S-0134219.
XX 14-MAY-1999; 990S-0134221.
XX 14-MAY-1999; 990S-0134370.
XX 18-MAY-1999; 990S-0134768.
XX 19-MAY-1999; 990S-0134941.
XX 20-MAY-1999; 990S-0135124.
XX 21-MAY-1999; 990S-0135453.
XX 24-MAY-1999; 990S-0135629.
XX 25-MAY-1999; 990S-0136021.
XX 27-MAY-1999; 990S-0136392.
XX 28-MAY-1999; 990S-0136782.
XX 01-JUN-1999; 990S-0137222.
XX 03-JUN-1999; 990S-0137528.
XX 04-JUN-1999; 990S-0137502.
XX 07-JUN-1999; 990S-0137724.
XX 08-JUN-1999; 990S-0138094.
XX 10-JUN-1999; 990S-0138540.
XX 10-JUN-1999; 990S-0138847.
XX 14-JUN-1999; 990S-0139119.
XX 16-JUN-1999; 990S-0139452.
XX 16-JUN-1999; 990S-0139453.
XX 17-JUN-1999; 990S-0139492.
XX 18-JUN-1999; 990S-0139454.
XX 18-JUN-1999; 990S-0139455.
XX 18-JUN-1999; 990S-0139456.
XX 18-JUN-1999; 990S-0139457.
XX 18-JUN-1999; 990S-0139458.
XX 18-JUN-1999; 990S-0139459.
XX 18-JUN-1999; 990S-0139460.
XX 18-JUN-1999; 990S-0139461.
XX 18-JUN-1999; 990S-0139462.
XX 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139764.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140454.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140824.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141297.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142590.
PR 08-JUL-1999; 990S-0142804.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 24-JUL-1999; 990S-0145224.
PR 25-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147204.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147933.
PR 10-AUG-1999; 990S-0148371.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.


```

PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151066
PR 27-AUG-1999; 99US-0151080
PR 30-AUG-1999; 99US-0151303
PR 31-AUG-1999; 99US-0151338
PR 01-SEP-1999; 99US-0151930
PR 07-SEP-1999; 99US-0152363
PR 10-SEP-1999; 99US-0153070
PR 13-SEP-1999; 99US-0153758
PR 15-SEP-1999; 99US-0154018
PR 16-SEP-1999; 99US-0154039
PR 20-SEP-1999; 99US-0154779
PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157753
PR 06-OCT-1999; 99US-0157865
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232
PR 12-OCT-1999; 99US-0158364
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159295
PR 14-OCT-1999; 99US-0159329
PR 14-OCT-1999; 99US-0159330
PR 14-OCT-1999; 99US-0159331
PR 14-OCT-1999; 99US-0159637
PR 14-OCT-1999; 99US-0159638
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741
PR 21-OCT-1999; 99US-0160767
PR 21-OCT-1999; 99US-0160768
PR 21-OCT-1999; 99US-0160770
PR 21-OCT-1999; 99US-0160814
PR 21-OCT-1999; 99US-0160815
PR 22-OCT-1999; 99US-0160980
PR 22-OCT-1999; 99US-0160981
PR 22-OCT-1999; 99US-0160989
PR 25-OCT-1999; 99US-0161404
PR 25-OCT-1999; 99US-0161405
PR 25-OCT-1999; 99US-0161406
PR 26-OCT-1999; 99US-0161359
PR 26-OCT-1999; 99US-0161360
PR 26-OCT-1999; 99US-0161361
PR 28-OCT-1999; 99US-0161920
PR 28-OCT-1999; 99US-0161942
PR 28-OCT-1999; 99US-0161943
PR 29-OCT-1999; 99US-0162142

Query Match
Best Local Similarity 37.68; Pred. No. 2, 1c 67;
Matches 184; Conservative 91; Mismatches 193; Indels 22; Gaps 3,

QY 9 MYDQCFVTR-----GDWID-----VYNPATEAVISRPDGOARTAP 47
DB 1 MSMDAQSVSEKRSGLRQGLIGSKWIGSYINKILKVNNEAFSGFIALVA-MRIKFN 90
QY 48 KAIDAAEPAPENALAIERASWIKTSACTIPERASEISALIVPRGKIQALAEVFAF 107
DB 61 DATASYEFTSWSPITAGERSKVLRFWYGLLIAHKEELQLITLQGRKREAGEVAY 120
QY 108 TADYIDYMAEWARYEGFTQSOPGDNILLERKALGVITIGILPWNFPFLAKKMAPAL 167
DB 121 GASFLYYAEAKKRYGDIIPPNLSDRLLVLKQPGVGTGATPPWNEPLAMITRKVSPAL 180
QY 168 LTRNTVTIKFEFTTNAIAFAKIVDFLCLPPGVFNVLGRGETVQDELQAGNFKVAMVSM 227
DB 181 ASGCTVVVKPSBELPLTALAAELAQAGVPPGALNVVGNAPELIGDALLTSPQVRKITF 240
QY 228 TCSVSA-FKIMATAAKNITKVCLCLOGLKAIJWEDADLELAVKAIVUSRVINSQVNM* 247

```

RESULT 13

AAC23559

ID AAC23559 standard; Protein; 493 AA.

XX AC AAC23559;

XX DI 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26912.

XX KW Protein identification, signal transduction pathway, metabolic pathway,

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2800EP-0301439

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123419.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 04-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 23-APR-1999; 99US-0130891.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 30-APR-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 04-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134758.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

PR 24 MAY-1999; 9905-0135629.
PR 25 MAY-1999; 9905-0136021.
PR 27 MAY-1999; 9905-0136392.
PR 28 MAY-1999; 9905-0136782.
PR 01 JUN-1999; 9905-0137222.
PR 03 JUN-1999; 9905-0137528.
PR 04 JUN-1999; 9905-0137502.
PR 07 JUN-1999; 9905-0137724.
PR 08 JUN-1999; 9905-0138094.
PR 10 JUN-1999; 9905-0138540.
PR 10 JUN-1999; 9905-0138847.
PR 14 JUN-1999; 9905-0139119.
PR 16 JUN-1999; 9905-0139452.
PR 16 JUN-1999; 9905-0139453.
PR 17 JUN-1999; 9905-0139492.
PR 18 JUN-1999; 9905-0139454.
PR 18 JUN-1999; 9905-0139455.
PR 18 JUN-1999; 9905-0139456.
PR 18 JUN-1999; 9905-0139457.
PR 18 JUN-1999; 9905-0139458.
PR 18 JUN-1999; 9905-0139459.
PR 18 JUN-1999; 9905-0139460.
PR 18 JUN-1999; 9905-0139461.
PR 18 JUN-1999; 9905-0139462.
PR 18 JUN-1999; 9905-0139463.
PR 18 JUN-1999; 9905-0139750.
PR 18 JUN-1999; 9905-0139763.
PR 21 JUN-1999; 9905-0139817.
PR 22 JUN-1999; 9905-0139899.
PR 23 JUN-1999; 9905-0140353.
PR 23 JUN-1999; 9905-0140354.
PR 24 JUN-1999; 9905-0140695.
PR 28 JUN-1999; 9905-0140823.
PR 29 JUN-1999; 9905-0140991.
PR 30 JUN-1999; 9905-0141287.
PR 01 JUL-1999; 9905-0141842.
PR 01 JUL-1999; 9905-0142154.
PR 02 JUL-1999; 9905-0142055.
PR 06 JUL-1999; 9905-0142390.
PR 08 JUL-1999; 9905-0142803.
PR 09 JUL-1999; 9905-0142920.
PR 12 JUL-1999; 9905-0142977.
PR 13 JUL-1999; 9905-0143542.
PR 14 JUL-1999; 9905-0143624.
PR 15 JUL-1999; 9905-0144005.
PR 16 JUL-1999; 9905-0144085.
PR 16 JUL-1999; 9905-0144086.
PR 19 JUL-1999; 9905-0144325.
PR 19 JUL-1999; 9905-0144331.
PR 19 JUL-1999; 9905-0144342.
PR 19 JUL-1999; 9905-0144343.
PR 19 JUL-1999; 9905-0144344.
PR 19 JUL-1999; 9905-0144345.
PR 20 JUL-1999; 9905-0144352.
PR 20 JUL-1999; 9905-0144632.
PR 20 JUL-1999; 9905-0144884.
PR 21 JUL-1999; 9905-0144814.
PR 21 JUL-1999; 9905-0145086.
PR 21 JUL-1999; 9905-0145088.
PR 22 JUL-1999; 9905-0145085.
PR 22 JUL-1999; 9905-0145087.
PR 22 JUL-1999; 9905-0145089.
PR 22 JUL-1999; 9905-0145192.
PR 23 JUL-1999; 9905-0145192.
PR 23 JUL-1999; 9905-0145192.
PR 23 JUL-1999; 9905-0145218.
PR 24 JUL-1999; 9905-0145224.
PR 24 JUL-1999; 9905-0145276.
PR 27 JUL-1999; 9905-0145913.
PR 27 JUL-1999; 9905-0145918.
PR 27 JUL-1999; 9905-0145919.
PR 28 JUL-1999; 9905-0145951.
PR 02 AUG-1999; 9905-0146386.
PR 02 AUG-1999; 9905-0146388.
PR 02 AUG-1999; 9905-0146489.
PR 03 AUG-1999; 9905-0147048.
PR 04 AUG-1999; 9905-0147204.
PR 04 AUG-1999; 9905-0147402.
PR 05 AUG-1999; 9905-0147192.
PR 05 AUG-1999; 9905-0147260.
PR 06 AUG-1999; 9905-0147403.
PR 06 AUG-1999; 9905-0147416.
PR 09 AUG-1999; 9905-0147494.
PR 09 AUG-1999; 9905-0147945.
PR 10 AUG-1999; 9905-0148171.
PR 11 AUG-1999; 9905-0148419.
PR 12 AUG-1999; 9905-0148441.
PR 13 AUG-1999; 9905-0148565.
PR 13 AUG-1999; 9905-0148684.
PR 16 AUG-1999; 9905-0149368.
PR 17 AUG-1999; 9905-0149175.
PR 18 AUG-1999; 9905-0149426.
PR 20 AUG-1999; 9905-0149722.
PR 20 AUG-1999; 9905-0149723.
PR 20 AUG-1999; 9905-0149929.
PR 23 AUG-1999; 9905-0149902.
PR 23 AUG-1999; 9905-0149930.
PR 25 AUG-1999; 9905-0150566.
PR 26 AUG-1999; 9905-0150884.
PR 27 AUG-1999; 9905-0151065.
PR 27 AUG-1999; 9905-0151066.
PR 27 AUG-1999; 9905-0151080.
PR 30 AUG-1999; 9905-0151303.
PR 31 AUG-1999; 9905-0151438.
PR 01 SEP-1999; 9905-0151930.
PR 07 SEP-1999; 9905-0152363.
PR 10 SEP-1999; 9905-0154070.
PR 13 SEP-1999; 9905-0154758.
PR 15 SEP-1999; 9905-0154018.
PR 16 SEP-1999; 9905-0154039.
PR 20 SEP-1999; 9905-0154779.
PR 22 SEP-1999; 9905-0155139.
PR 23 SEP-1999; 9905-0155486.
PR 24 SEP-1999; 9905-0155659.
PR 28 SEP-1999; 9905-0156458.
PR 29 SEP-1999; 9905-0156596.
PR 04 OCT-1999; 9905-0157117.
PR 05 OCT-1999; 9905-0157753.
PR 06 OCT-1999; 9905-0157865.
PR 07 OCT-1999; 9905-0158029.
PR 08 OCT-1999; 9905-0158232.
PR 12 OCT-1999; 9905-0158369.
PR 13 OCT-1999; 9905-0159294.
PR 13 OCT-1999; 9905-0159294.
PR 14 OCT-1999; 9905-0159295.
PR 14 OCT-1999; 9905-0159429.
PR 14 OCT-1999; 9905-0159430.
PR 14 OCT-1999; 9905-0159441.
PR 14 OCT-1999; 9905-0159637.
PR 14 OCT-1999; 9905-0159648.
PR 18 OCT-1999; 9905-0159584.
PR 21 OCT-1999; 9905-0160741.
PR 21 OCT-1999; 9905-0160767.
PR 21 OCT-1999; 9905-0160768.
PR 21 OCT-1999; 9905-0160770.
PR 21 OCT-1999; 9905-0160814.
PR 21 OCT-1999; 9905-0160815.
PR 22 OCT-1999; 9905-0160980.
PR 22 OCT-1999; 9905-0160981.
PR 22 OCT-1999; 9905-0160989.
PR 25 OCT-1999; 9905-0161404.
PR 25 OCT-1999; 9905-0161405.
PR 25 OCT-1999; 9905-0161406.
PR 26 OCT-1999; 9905-0161559.
PR 26 OCT-1999; 9905-0161660.
PR 26 OCT-1999; 9905-0161861.
PR 28 OCT-1999; 9905-0161920.

```

PP 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 33.7%, Score 825.5; DB 21, Length 493;
Best Local Similarity 38.6%, Pred No 2 3e-67,
Matches 180; Conservative 88; Mismatches 189; Indels 9; Gaps 2;

QY 20 GDAWID-----VVFATAVSKRIPDQAEIAAKKALUAARAKAUEWAEKALVAIGRWSW 71
Db 23 GKKWIDYNDKTIKVNPNATGEIADVACMGTKEDINDAIASVEAFSTSKRTASGRSV 82
QY 72 LRFISAGIRPASFSIALIVEGSKIQQLAEVEVAFADYIDVMAEWAPYEGEITQSDR 131
Db 83 LRRWYDLLTAHKEELGQIILEGCKPLKEALGEVAGVAGSPLEYAEAKVYDUIPPNL 142
QY 132 PGNILLKRALGVTTGILPNWFFFLIARKMAPALTGNTIVIKSEFTNNAIAFAKI 191
Db 143 SDFRTIVIKQVGVIGVATFWNFELAMITRKVGIFALASCTVWVKTSILPIALAAEL 202
QY 192 VDEIGLPRGVNLVGRGTYGQELAGNPKVAMVSMGTGVSAGEKIMATAAKNITKVCLE 251
Db 203 ALGAGVPGALNVWGNNAFELGALLTSPVKEIIFLGS:AVGKKLMAAANP1VKKVSLB 262
QY 252 LGKAPATIMDDADELAVRAIVDSRVINSQVNCNCAEPYVYVKGIGYDQVNP1PAMQA 311
Db 263 LGGNAPSIVLGLALLVAVKGIILAAKFRNSJQTCVANKVLVGLGIDYKFAEAFSEAVCK 322
QY 312 VQFGNPAERNDIAMPILNAAALERVQKVARAVEGARGVAFGKVGKGYYPPTLLL 371
Db 323 LEVGD-GRDGTGGLINDAAQVQVETVQDAVSKGAKIIGGKRHSGLGMTYEPTVTR 381
QY 372 DVQRQMSIMHETGPGVPLVAVFDTLEDAISMANDSDYGLTSSIIYUNLVAMKALGKLG 431
Db 382 DVSDNWKSEIEFGVAPLIRPKTEEDAIRANDTIAGLAAYIFTNSVORSNRVFEALE 441
QY 432 PGETVINREPMGGFIAGKWKSGICGADCKKHGICHOYIQTQVYVL 477
Db 442 YGLVAVNEGLISTEVAPPRVYKQSLGPGFYSKYGMDEYLEIKYVCL 487

RESULT 14
AB447807
ID AB447807 standard; Protein: 488 AA
AC
XX
AC
XX
UT
XX
DE
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes protein #511.
XX
PN Listeria monocytogenes protein #511.
XX
PD WO200177335-A2.
XX
PP 18-OCT-2001.
XX
PP 11-APP-2001; 2001WO-PP01118
XX
PP 11-APP-2000; 2000EP-0004620.
XX
PA (INSP ) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetolani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia F, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

```

```

FI Rose M, Voss H,
XX WFL 2002-010914/01.
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides
XX
XX Claim 6; SEQ ID No 512; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccines compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 488 AA:
XX
XX Query Match 33.2%, Score 811.5; DB 23; Length 488;
XX Best Local Similarity 46.4%, Pred No 4 4e-66;
XX Matches 175; Conservative 107; Mismatches 186; Indels 13; Gaps 4;
XX
QY 5 VOHPMYIDQVTVRGDAWIDVNPATEAVLSRIPDQGAEDAKKADAARAPPEWALP 64
Db 11 VQTKIFNGKWTGINKETKIDVNPANVDVTAQAGPSEFKATKAATAPDPAKMF 79
QY 65 ATEPASWPKTSAGTPTFPASRISALIVFEGGKIQQLAEVEVAFADYIDVMAEWAPYFG 124
Db 71 LADRVKLLKHLADLMEEKADTLAKIMILEQCKPLKESKGEVLTCVENFPAEAEARLYG 130
QY 125 EIIQSDPPGEN---IIFKPAIGVTTGILPNWFFFLIARKMAPALTGNTIVIKSEFT 181
Db 131 ETI-----PAPNNHAFIVKKQPIGVVAAITPWNFPGCMVTRKLAIPALATNTIVLKPSGDI 186
QY 182 TNNATAFKIVDFEICLPQGVENIVLGRGTYGQELAGNPKVAMVSMGTGVSAGEKIMATA 241
Db 187 PLSAIAIPEIFEEASLQKGVANIVMGSSEKTEGRIITISLAVPKIIFIRSKVQGITPKQS 246
QY 242 AKNIITKVLELGSKKAIATVMEALGLELAVRAIVLSRVINSQVNCNCAEPYVYVKGIGYDQF 401
Db 247 ADTLKKISLELGGHAPFIVFDADNDAANLVAAKPKNNGCVCVSPNRTFVAKETKEKF 406
QY 302 VNP1CEAMQAVGFONPAEPNDIAMSPLINAAALERVQKVARAVEGARGVAFGKVGKAVG- 460
Db 307 TKALVAKVEQLKVGNGLD--DVNVVGPILIPFAIUKIDKQLKNATEKCAKVIITGGRPLTS 464
QY 361 ---KGYVYPPTLTLDVPCQMSIMHETGPGVPLVAVFDTLEDAISMANDSDYGLTSSIIY 417
Db 365 DYDKGNFYKPTVLNDVTKKMDIFYEETFGVPIPLITFETEDALEMANSEGLASYFYT 424
QY 418 QNLNVAMKAIKSLKPRFIVINPNEFAMQGFHAWRKSRSGGAAKRGHLSHGYIQTQVYVL 477
Db 425 KDLAPVERVGAALVGVGMVGAELIASNPPTPGCVKHSQGPFGNGHYGMFETQVRFINL 484
QY 478 Q 478
Db 485 K 485

```

RESULT 15

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:13:55, Search time 13.2654 seconds
(without alignments)
1062.445 Million cell updates/sec

Title: US-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVPVQHPMYIDQGFVTKG, AAKKHSIHGYLFGVVYIQS 474

Scoring table: BLOSUM62

Gapop 10 0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5H_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	932.5	38.1	493	4	US-09-134-001C-4388	Sequence 4388, Ap
2	767.5	31.4	506	4	US-09-134-001C-4383	Sequence 4383, Ap
3	737.5	30.2	486	4	US-09-134-001C-4246	Sequence 4246, Ap
4	683.5	27.9	497	2	US-08-513-841-2	Sequence 2, Appl
5	683.5	27.9	497	2	US-08-696-834-2	Sequence 2, Appl
6	683.5	27.9	497	2	US-08-942-634-2	Sequence 2, Appl
7	683.5	27.9	497	4	US-09-118-317-2	Sequence 2, Appl
8	669.5	27.4	521	4	US-09-231-294-2	Sequence 2, Appl
9	639.5	26.1	482	4	US-09-155-183-4	Sequence 4, Appl
10	633.5	25.9	510	4	US-09-134-001C-4541	Sequence 4541, Ap
11	625.5	25.6	518	4	US-09-134-001C-4451	Sequence 4451, Ap
12	623.5	25.5	486	4	US-09-031-941-2	Sequence 9, Appl
13	623.5	25.5	486	4	US-09-055-597-2	Sequence 9, Appl
14	623.5	25.5	508	4	US-09-055-2708-4	Sequence 9, Appl
15	448	18.3	487	4	US-09-351-2248-5	Sequence 5, Appl
16	353.5	14.5	464	4	US-09-134-001C-4701	Sequence 4701, Ap
17	221.5	9.1	133	1	US-08-446-611-2	Sequence 2, Appl
18	221.5	9.1	133	2	US-08-794-494-2	Sequence 2, Appl
19	156.5	6.4	551	4	US-08-952-061-2	Sequence 2, Appl
20	120.5	4.9	711	4	US-09-134-001C-5460	Sequence 18, Appl
21	119	4.9	3724	2	US-08-804-2270-10	Sequence 4, Appl
22	119	4.9	3724	2	US-08-804-198-4	Sequence 4, Appl
23	118	4.8	72	4	US-08-655-2708-33	Sequence 30, Appl
24	101.5	4.1	1704	4	US-08-485-355B-40	Sequence 40, Appl
25	100	4.1	564	1	US-08-427-097-2	Sequence 2, Appl
26	100	4.1	564	1	US-08-427-097-2	Sequence 16, Appl
27	100	4.1	564	2	US-08-878-957-2	Sequence 2, Appl

28	100	4.1	564	2	US-08-878-957-16	Sequence 16, Appl
29	99.5	4.1	722	2	US-08-843-530B-28	Sequence 28, Appl
30	99.5	4.1	1281	2	US-08-843-530B-6	Sequence 6, Appl
31	99.5	4.1	1298	2	US-08-843-530B-2	Sequence 2, Appl
32	99.5	4.1	1298	2	US-08-843-530B-4	Sequence 4, Appl
33	99.5	4.1	1298	2	US-08-843-530B-44	Sequence 34, Appl
34	99.5	4.1	2285	4	US-09-308-375-2	Sequence 2, Appl
35	99	4.0	564	1	US-08-427-097-14	Sequence 14, Appl
36	94	4.0	564	1	US-08-427-097-20	Sequence 20, Appl
37	94	4.0	564	1	US-08-878-957-14	Sequence 14, Appl
38	99	4.0	564	2	US-08-878-957-20	Sequence 20, Appl
39	99	4.0	564	2	US-08-878-957-28	Sequence 28, Appl
40	99	4.0	564	2	US-07-603-133B-12	Sequence 12, Appl
41	98	4.0	775	2	US-08-878-957-22	Sequence 22, Appl
42	96.5	3.9	564	2	US-08-853-659A-51	Sequence 51, Appl
43	96.5	3.9	1289	2	US-09-012-030-2	Sequence 2, Appl
44	96.5	3.9	743	2	US-08-590-451-2	Sequence 2, Appl
45	95.5	3.9	743	2	US-08-590-451-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-4388
; Sequence 4388, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4388
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4388

Query Match	38.1%	Score 932.5;	DB 4;	Length 493;
Best Local Similarity	38.5%	Pred. No. 7.4e-88;		
Matches	181,	Conservative 108,	Mismatches 180,	Indels 1; Gaps 1;
QY	9	MYHAGQFVWEGHAWHVVNPATFAVLSRIPIQAGAFELAKKATLUAAPAPGPEWALPATER	68	
DB	24	LFNNETESQSYETIMVTPATCFATTTTATFEFVHIVATKESQCAQIWEFVQPTF	83	
QY	69	ASWQKFSAGIPFAEISALIVEEGKQGLAEVIVALTATYLYMAWAPFYDEI	128	
DB	84	AFHKLLIPLEKNPKQFAGLYVKBQKTLIAAYGHIKSIPTFIMISLSWSGSRVQ	143	
QY	129	SDRPSNELLFKVALGVITGILPWNPFPLAPKMAPALLTGNITVIFSEPTNNIAF	188	
DB	144	NSIANETLQINKFGVAGIVVNAIILVLMKVIYFAIVDSVVKESLGLILLG	204	
QY	189	AKIVGEZLPGVENVLDSGEIVGQELACNKKVAMVSMIGVSAGFFIMAAIAFNIRV	248	
DB	264	NLEGGNAVIVTSSNAELKAVNYVIAPIINNAQVCTCFEYFVHESVHEIFLNKVTSK	324	
QY	309	MQAVQGNPAFPNDJAMGPPLINAAALFFKPKQKVARAVRFAVAGKAVEGKGYYPPT	368	
DB	324	MKSLVGVGFLEN-TLYGAILNKQLLSITHEKVQDAIKNCATIMTSGHLKPKGFFYAPT	382	


```

1  ZIP 22202
2  COMPUTER READABLE FORM:
3  MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
4  COMPUTER: IBM PC compatible
5  OPERATING SYSTEM: PC-DOS/MS-DOS
6  SOFTWARE: MS-DOS Editor
7  CURRENT APPLICATION DATA:
8  APPLICATION NUMBER: US/08/513,841
9  FILING DATE: 01-NOV-1995
10 CLASSIFICATION: 435
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: UK 9304700.9
13 FILING DATE: 08-MAR-1993
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: JP 241851/1993
16 FILING DATE: 28-SEP-1993
17 ATTORNEY/AGENT INFORMATION:
18 NAME: NORMAN F. ORLON
19 REGISTRATION NUMBER: 24,618
20 REFERENCE/DOCKET NUMBER: 1R-909-0 PCT
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: 703-413-3000
23 TELEFAX: 703-413-3220
24 INFORMATION FOR SEQ ID NO. 2:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 497 amino acids
27 TYPE: amino acid
28 TOPOLOGY: linear
29 MOLECULE TYPE: peptide
30 ORIGINAL SOURCE:
31 ORGANISM: Gluconobacter oxydans
32 STRAIN: T-100
33 FEATURE:
34 NAME/KEY: mat peptide
35 LOCATION: 1..497
36 IDENTIFICATION METHOD: experimentally
37 US-08-513-841-2

```

```

Query Match 27.9%, Score 683.5, DB 1: Length 497;
Best Local Similarity 33.1%, Pred. No. 56-62;
Matches 159; Conservative 107; Mismatches 206; Indels 9; Gaps 6;

QY 4 PVQHPMYIDGQFVWVR-GDAWIIUVNPAFAVISPPIGAGAEAKAALIAALKA--QPPW 60
DB 13 PRRGFFIDGR--WRACKDFDFORSSPAHVPVPTPPCTRESLDEAVAAAPAFENGSW 69
QY 61 EALPAIEPASWLRKISAGIPEPASEISALIVEGGKICVLAIEVEVAFTADYIIVNAEWAP 120
DB 70 AGLAAADRAAVALIKAAGLIRERDDIAYWEVLENGKPTISQAKGEIDHCTACPEMAAGAP 129
QY 121 PYEGEITQSGRRPGENTILFKPALGVITGILPWNEFFELIARFMAPALLTNTIVIKSPFF 180
DB 130 MLCGTENNIGRGGLFGCVLREPIGVVGLITPWNPPMILCPAPPFIIASGCTLVKPAEV 189
QY 181 TTNNIAFAKIVIVETILPPQVPLVIGAPPTVGLAGNPKVAMVMTGSVAGKK-IMA 239
DB 190 TSATITLLAEIADAGLPGKGVNVVITGTGTVCQAMTEHQIDMLSTGTSGVGSCTHA 249
QY 240 TAANKITKVGLEGGKAPAVLMDADLELAVKAIVSPVINSQVNCAPPVYVQKGIYD 299
DB 250 AADSNLKKLGLEGGKRPVVFADSNLEDAADAVAFGISFNTGQCCVSSRLIVERVAE 309
QY 309 GIVNFIQEMAVGIFPPFAENFTIAMFTTHAAALFEVFGKVAFAVEGAFVAFGKAVF 369
DB 310 KFERLVVPEKMEKIRVGPPD-PEIQGATTEAGNKIILDYIAKKAEGAKLLCGGIVD 368
QY 360 -GKGYYPPTLLLDVREMSIMHEETGGVPLVPVAFDTLLEDATISMANDSYGLTSIYTO 418
DB 369 PKKQYIQTPTITVIVKFSMGIAPQELGEPVLAISFHFIDVDEALAIANDVYVGLAASVMSK 428
QY 419 NINVAMKAIGLKGFTYINIRENFEAMOGFHAGKWSKIGGADGKHLGHLYLOTQVYILO 478

```

```

DB 429 DICKALAVTRPVAGPEWNTIMSGRETHGFKSOWEPACLYGVVEYTGIFSVHIE 488
QY 479 S 479
DB 489 T 489

RESULT 5
US-08-696-834-2
; Sequence 2, Application US/08696834
; Patent No. 5834263
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Hayashi, Hiromi
; TITLE OF INVENTION: Method for Producing 2-Keto-L-Gulonic Acid
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ohlon, Spivak, McLeiland, Maier & Neustadt,
; ADDRESS: P.O.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,834
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 28612/1994
; FILING DATE: 25-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-3220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Gluconobacter oxydans
; STRAIN: T-100
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 1..497
; IDENTIFICATION METHOD: experimentally
; US-08-696-834-2

```

```

Query Match 27.9%, Score 683.5, DB 2: Length 497;
Best Local Similarity 33.1%, Pred. No. 56-62;
Matches 159; Conservative 107; Mismatches 206; Indels 9; Gaps 6;

QY 4 PVQHPMYIDGQFVWVR-GDAWIIUVNPAFAVISPPIGAGAEAKAALIAALKA--QPPW 50
DB 13 PRRGFFIDGR--WRACKDFDFORSSPAHVPVPTPPCTRESLDEAVAAAPAFENGSW 69
QY 61 EALPAIEPASWLRKISAGIPEPASEISALIVEGGKICVLAIEVEVAFTADYIIVNAEWAP 120
DB 70 AGLAAADRAAVALIKAAGLIRERDDIAYWEVLENGKPTISQAKGEIDHCTACPEMAAGAP 129

```

QY 121 FYEGELIUSDRPGENIIIEKRALGVTTGILPWNEPFLIARKKAPAILTGNTIVIKPSEF 180
 Db 130 MHRDFTNNIGELGFWVI PFPITGVVGLITPWNFPMTITPPAPFILASGTTLVVKPAEV 189
 QY 181 TTNATAFAKIVDEIGLPGVNLVIGRGETVGOELAGNPKVAMVSMTGVSAGKE-IMA 239
 Db 190 TSATILLAILADAGLPKGVNVVTGTRTCQAMTEHODIDMLSTGTCVCKSCIIA 249
 QY 240 TAAKNITKVCLEGGKAPALVMDADLELAVKAVDSRVINSQVCNCAERVVYKGIYD 299
 Db 250 AADSNLKKIGLEGGKNPIVFAADSNLEDAADAVAFISFNTGCCVSSRLIVERSVAE 309
 QY 300 QFVNRLEGAQVQFQNPARNDIAMGPLINAAALERVEQKVARAVEGCAVAFGKAVE 359
 Db 310 KFERLVVPKMEKIRVGDPEP-PEIQIGCAITTTAQNKTTLDYIAKKAFCALLTGCGIYD 369
 QY 360 -GKGYVYPPTLLDVOKQMSIMHEETFGVPLVPWAFDTLEDALSMANDSDYGLTSIYVQ 418
 Db 369 FKGQYIQTPLFTIVKPSMGIAHDFIFGVLASFHFDVDEAIAIANDTVYGLAASVWSK 428
 QY 419 NLNVAMKAIKGLAFGETYINRENFEAMOGFHAGWKSGLGCGADCKKHGLIGLYLOVYVLO 478
 Db 429 DIFKALAVTRVRVAGRPWNTINSQGPETPLGPKQSGMRAGLYGVVEEYTOIKSVHIE 488
 QY 479 S 479
 Db 489 T 489

RESULT 6

US-08-942-673-2
 ; Sequence 2, Application US/08942673
 ; Patent No. 5861292
 ; GENERAL INFORMATION:
 ; APPLICANT: Niwa, Mineo
 ; APPLICANT: Saito, Yoshimasa
 ; APPLICANT: Ishii, Yoshinori
 ; APPLICANT: Yoshida, Masaru
 ; APPLICANT: Suzuki, Hiromi
 ; TITLE OF INVENTION: No. 5861292-i L-sorbose dehydrogenase and No. 5861292-i
 ; TITLE OF INVENTION: L-sorbose dehydrogenase obtained from Gluconobacter
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Obolon, Spivak, McClelland, Maier & Neustadt, P.C.
 ; STREET: 1755 Jefferson Davis Highway, Suite 400
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette 3.50 inch, 1.44 Mb storage
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MS-DOS Editor
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: US/08/942,673
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: 08/513,841
 ; FILING DATE: 01-NOV-1995
 ; APPLICATION NUMBER: HK 9404700 9
 ; FILING DATE: 08-MAR-1993
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: JP 241851/1993
 ; FILING DATE: 28-SEP-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: NORMAN F. OBLON
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 18-909-0 PCT
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 703 413-4000
 TELEFAX: 703 413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 ORGANISM: Gluconobacter oxydans
 STRAIN: T-100
 FEATURE:
 NAME/KEY: mat peptide
 LOCATION: 1-497
 IDENTIFICATION METHOD: experimental y
 US-08-942-673-2

Query Match 27.98; Score 681.6; DB 2; Length 497;
 Best local Similarity 33.18; Pred. No. 5861292;
 Matches 159; Conservative 107; Mismatches 206; Indels 9; Gaps 6;
 QY 4 PVCHPMYIDGQFVTPW-GLAWIIVVNVNATEAVISLPIRQQAIAIAKAAHAAERA QYEW 60
 Db 13 PREFGFEIDCE--WFAKKEDEFKRSFAHGVIVIPRITREIDDEAVAAAKRAELNGSW 69
 QY 61 EALPATERASWLKIKISAGIKERKASELSALIVPECKIIOULAFVFAVFAIDVITYMAEMAR 120
 Db 70 AGLAAARAAVLLKAAGLLFERRDDIAYWEVLENKPIISQAKGELIDP IATFMAAAAR 129
 QY 121 FYEGELIUSDRPGENIIIEKRALGVTTGILPWNEPFLIARKKAPAILTGNTIVIKPSEF 180
 Db 130 MLHGDTFNNGELGFWVLEKPIGVVGLITPWNFPMTITPPAPFILASGTTLVVKPAEV 189
 QY 181 TTNATAFAKIVDEIGLPGVNLVIGRGETVGOELAGNPKVAMVSMTGVSAGKE-IMA 239
 Db 190 TSATILLAILADAGLPKGVNVVTGTRTCQAMTEHODIDMLSTGTCVCKSCIIA 249
 QY 240 TAAKNITKVCLEGGKAPALVMDADLELAVKAVDSRVINSQVCNCAERVVYKGIYD 299
 Db 250 AADSNLKKIGLEGGKNPIVFAADSNLEDAADAVAFISFNTGCCVSSRLIVERSVAE 309
 QY 300 QFVNRLEGAQVQFQNPARNDIAMGPLINAAALERVEQKVARAVEGCAVAFGKAVE 359
 Db 310 KFERLVVPKMEKIRVGDPEP-PEIQIGCAITTTAQNKTTLDYIAKKAFCALLTGCGIYD 369
 QY 360 -GKGYVYPPTLLDVOKQMSIMHEETFGVPLVPWAFDTLEDALSMANDSDYGLTSIYVQ 418
 Db 369 FKGQYIQTPLFTIVKPSMGIAHDFIFGVLASFHFDVDEAIAIANDTVYGLAASVWSK 428
 QY 419 NLNVAMKAIKGLAFGETYINRENFEAMOGFHAGWKSGLGCGADCKKHGLIGLYLOVYVLO 478
 Db 429 DIFKALAVTRVRVAGRPWNTINSQGPETPLGPKQSGMRAGLYGVVEEYTOIKSVHIE 488
 QY 479 S 479
 Db 489 T 489

RESULT 7

US-09-118-317-2
 ; Sequence 2, Application US/09118417
 ; Patent No. 6197562
 ; GENERAL INFORMATION:
 ; APPLICANT: Niwa, Mineo
 ; APPLICANT: Saito, Yoshimasa
 ; APPLICANT: Ishii, Yoshinori
 ; APPLICANT: Yoshida, Masaru
 ; APPLICANT: Suzuki, Hiromi
 ; TITLE OF INVENTION: No. 6197562-i L-sorbose dehydrogenase and No. 6197562-i
 ; TITLE OF INVENTION: L-sorbose dehydrogenase obtained from Gluconobacter
 ; TITLE OF INVENTION: oxydans T-100
 ; NUMBER OF SEQUENCES: 22


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Oshon, Spivak, McClelland, Maier & Neustadt, P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1 44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS Editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,317
FILING DATE:
CLASSIFICATION:
PROIP APPLICATION DATA:
APPLICATION NUMBER: 08/513,841
FILING DATE: 01-NOV-1995
APPLICATION NUMBER: UK 9304700 9
FILING DATE: 08-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 241851/1993
FILING DATE: 28 SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-909-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Gluconobacter oxydans
STRAIN: T-100
FEATURE:
NAME/KEY: mat peptide
LOCATION: 1..497
IDENTIFICATION METHOD: expected neutrality
US-09-118-317-2

Query Match 27 28, Score 683.5, DB 4, Length 497.
Best Local Similarity 33.19, Freq. No. 50-62.
Matches 159; Conservative 107; Mismatches 206; Indels 9; Gaps 6;

QY 4 PVQHPMYIDGQFVTK-GHAWILVGNKATEAVISRFIDGAAELAKKALDAEKA--QPEW 60
DB 13 PREGFFIDGE---WRAGKDFDFRSPAHNVPTPIPTCPDDEAVAAAPAFENSW 69
QY 61 EALPAIEPASMRTSANGTPERASPTSAITVERGKIQGLAEVEVAFTADYIDYMAEWA 120
DB 70 AGLAAADRAAVILAKAGLLBRDDIAYWEVLENGKPTISQAKGEIDHCITACFEMAAGA 129
QY 121 RYGEITQSDPPGRENIIIFKPAIGVTTGILPNWPFELIAPKMAPALLTGNITVIKPS 180
DB 130 MLHGDTNNLGEGLFGMLVREPIGVGLITPNWPFELIAPKMAPALLTGNITVIKPS 189
QY 181 TTNNAIAFAKIVIDEIGLPGGVFNVLVGRGETVQGLAGNPKVMVSMIGSVSAGEK-IMA 239
DB 190 TSAITLLLAELADAGLPKGVENVVGTGRTVQAMTEHGDIDMLSTFGTGVGKSCIIIA 249
QY 240 TAANKITKVLGIAGKAPATVMPDADLELAIVKIVSPVINSQVQNCNCAERYVYKGIYD 299
DB 250 AADSNLKKIGLIGLGGKNPVIVFAUSNLELAAGAVAFGLISFNIGQCCVSSSRILIVER 309
QY 300 QFVNPIGEAMCAVQFNPAPKKNIDAM3PLINAAALEKVEKLVKAPVAFKARVAFSGAVE 359
DB 359

310 KFEELVVFMEKLEFVDEFFS-TEFQGAIIIEA2KRIILLYLAKKAEAKYLLDGGIVD 368
369 GR3YVYFPIELLOVPEMSIMHEFTG3VILVVAHGLLELALSMAHSEYGLISSIVTQ 418
369 FGRGQYLCPTLFTLVKPKSMGIAHPTFEGVPLASPHPTVIEAIAIANDTVYSLAASV 428
369 FGRGQYLCPTLFTLVKPKSMGIAHPTFEGVPLASPHPTVIEAIAIANDTVYSLAASV 428
419 NLNVAMKAIKSLKEGETIYNRENFEAMOGFHAGWKSGLGGADGKHGLUGLYLOTQVV 478
419 NLNVAMKAIKSLKEGETIYNRENFEAMOGFHAGWKSGLGGADGKHGLUGLYLOTQVV 478
429 IIRKAIATVPRVPA3PFWNTIMSR3PFT4I33PKGS3W3P3AGLYGVBEYTOIKSVH 488
429 IIRKAIATVPRVPA3PFWNTIMSR3PFT4I33PKGS3W3P3AGLYGVBEYTOIKSVH 488
479 S 479
489 T 489

RESULT 8
US-09-221-294-2
Sequence 2, Application US/09221294
Patent No. 6268138
GENERAL INFORMATION:
APPLICANT: Riccardo Dalla-Favera and
APPLICANT: Alessandro Massimo Gianni
TITLE OF INVENTION: A Putt-vital V-actor Capable of Transducing the
TITLE OF INVENTION: Aldehyde Dehydrogenase-1 Gene and Uses of Said
TITLE OF INVENTION: Vector
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,294
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 42990-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-221-294-2

Query Match 27.48, Score 669.5, DB 4, Length 521.
Best Local Similarity 34.48, Freq. No. 150-60.
Matches 159; Conservative 89; Mismatches 197; Indels 17; Gaps 9;

QY 26 VWFATFAVIFPIHMGASHAPKAI-AALFA QFWEALAIAPASWIFKISASIPER 82
DB 56 VFNPAEEELQVFEETKFTVDKAVKAAQCAFQGTGSPWPTMPASEFGLLYKL-ADLIEP 114
QY 83 ASPLSALIVE---GKFLGLAHV-EVATTAIVIDYMAHWAFFYDE---ILQSDRGENT 136
DB 115 DRLLATMESMNGKLYSNAYINDLAGCIKTRFCAGWADKIOGGKITPID---GNF 171
QY 137 ILFKP--ALGVITGILPFPFFELIAFKMAPALLTGNITVIKPSFTTNNAIAFAKIV 194
DB 172 FTYTRHEPICVCGITITWNTFVLMILWKIGFALSCGNTVVVKFAGETFLIALHVASLI 231
QY 195 L3LPPQVFNVLGSGRIVG3ELASNPKNVAVMSLGSVSAEKLMAIAAK-NITKVLELS 253

```

```

Db 232 AGPPGVNIVPGYGTAGAAISSHMDIDKVAFTSTGVGKLIKKAAGSKLKRVTLEIG 291
Qy 254 GKAPAVMDADILEAVKALVDSRVINSQVCNCAERVVVGKGIYDQFVNRLGEMAOAVQ 313
Db 292 GKSPICVLADADNAVEFAHGVYHOGQCCIAASRI FVEBS IYDEFVRSVERAKKYI 351
Qy 314 FQPAENRDIAMGPIINAAALBRVOKVARAVEGKAVFGGKAVGGKGYPPPTLLLDV 373
Db 352 LGNPLTPG-VTQGPQDKQYKTIIDTFSKKGKKAUKLFGGCGPWCNKGQYFVQPTFSNV 410
Qy 374 RQMSIMHETFGVLPVVAFTLEIAIASMANUSYUGLTSIIYIIONLVANKAIKGLKFG 433
Db 411 TDMRIAKERIFGPVQQIMKPKSIDIVIKRANNTFYGLSAGVFTKIDKAITISSALQAG 470
Qy 434 ETYNKENEAMUGFHAGWKSGIGGAGCKHGLHGYLOTQVV 475
Db 471 TVWVNCYGVVSAQCPRGGFKMSONGRELGEYGFHEYTEVKT 512

RESULT 9
US-09-155-183-4
; Sequence 4, Application US/09155183
; Patent No. 6323011
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,183
; EARLIER FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

```

Query Match 26.1%; Score 630.5; DB 4; Length 482
Best Local Similarity 33.1%; Pred. No. 1.7e-57;
Matches 159; Conservative 88; Mismatches 216; Indels 17; Gaps 7;

Qy 8 PMYIDQGFVTRWTAMITVVNIPATFAVTSRIPGSGAFIAPKATUAATFAQFWEALIAIF 47
Db 5 PLLIGQSGPARKGRTPFRPNBVTGFLVSPVAAATLEDAADAAVAAQAQAFPAWALAPNE 64
Qy 68 RASWLRKISAGIPERASRISALITVEGSKTQOLAF---VEVAFTADYIIVYMAFWAPPEQ 124
Db 65 PRSEKIKAAQIQARSGE----FIFAAGETGAMAWYGFNRLAANMLREASMTTQVNC 129
Qy 125 ETQSDRPGENILKRALGVTTGILPWNPFFELIARKMAPALITGNITVIKPSFEFTINN 184
Db 121 EVIPSDVPGSFAMALRQPGVGLGIAPWNPVILATRAIAMPLACGNTVLKASELSFV 180
Qy 185 AIAFAKIVDFIGLPRGVNVLRCGFETVQ---ELAGNPKVAMVSMYSGVSAGKINATA 241
Db 181 HPIIGQVIGVAGHGGVWVNTSNAPADAQIVFRLIANPVPNPGFSTHVPFIVGELS 249
Qy 242 AKHITKVCLEGGKAPALIMDADILEAVKALVDSRVINSQVCNCAERVVVGKGIYDQFVN 301
Db 241 ARHLKPAILELGGKAPALIVDDADLEAAVQAAAFYNOGICMSTERLILVIAKADAP 300
Qy 302 VNLGEMAOVQFGNPAENRDIAMGPIINAAALBRVOKVARAVEGKAVFGGKAVGK 361
Db 301 VAGIAAKVETIRAGDPAHPESV-LGSLVDASAGTRIKALIDCAVAKGARILVIGQGL---E 356

```

```

Qy 362 GYYPTPTLLDVQKQMSIMHETFGVLPVVAFTLEIAIASMANUSYUGLTSIIYIIONLV 421
Db 357 GSTLOPTLLDGVASMBLYRRESPGIVAVVLGEGEALQIOTANISERHSAAITSRDGI 416
Qy 422 VANKAKIKKFEFTYINENF FAMQGPFAWRKSGIOMALGKRLHRYITQIVVYVQS 479
Db 417 KALALQORVSGLICHINGTVHREAMPF-GVKSQSGSPGKASTHEFTQIRWVILON 475

RESULT 10
US-09-134-001C-4541
; Sequence 4541, Application US/09144001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/144,001C
; CURRENT FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4541
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4541

```

```

Query Match 25.9%; Score 633.5; DB 4; Length 510;
Best Local Similarity 30.6%; Pred. No. 4e-57;
Matches 146; Conservative 100; Mismatches 222; Indels 9; Gaps 6;

Qy 9 MYIDQGFVTRWTAMITVVNIPATFAVTSRIPGSGAFIAPKATUAATFAQFWEALIAIF 68
Db 29 LFINPEQASDSGETLIVSNIPANDEIAKAPAKIKKFEFTYINENF FAMQGPFAWRKSGIOM 88
Qy 69 ASWIKPKISAGIPERASRISALITVEGSKTQOLAF---VEVAFTADYIIVYMAFWAPPEQ 127
Db 89 ADYLETSKRKHETHEIAIVDSGNGKPYRKESTIIVYCAANQKYFASVLTIDRESVN 148
Qy 128 QSDRPGENILKRALGVTTGILPWNPFFELIARKMAPALITGNITVIKPSFEFTINNATA 187
Db 149 FPIQGTMSLVVNVNFWVWGVVVAWNRHITLAKWKLKELIAAANFVVGQVTSSTWISLE 209
Qy 188 FAKLVIEIGLPRGVNVLRCGFETVQ---ELAGNPKVAMVSMYSGVSAGKINATA 247
Db 208 LAKIPGVVTPRGVNVNIDGSPSGDAIHHNVPKISPSGSLIVYVVAQAPGVIV 266
Qy 248 VTELEGGKAPALIMDADILEAVKALVDSRVINSQVCNCAERVVVGKGIYDQFVNRLG 307
Db 267 TTELEGGFSANITHEIANIPLVYHGVGQVGLIAPWNPVILATRAIAMPLACGNTVLK 426
Qy 308 AMCAVGVNPAENRDIAMGPIINAAALBRVOKVARAVEGKAVFGGKAVGK 362
Db 327 AFENIKVDFET-EFTKNSAGTGTGELQIFESYKILAEETGFANETLQAHPLIINGL 385
Qy 363 YYPPTPTLLDVQKQMSIMHETFGVLPVVAFTLEIAIASMANUSYUGLTSIIYIIONLV 422
Db 386 YFPEFTITENKIKIAGCEFTIPVVVVFVRETFQRALEIANISYGLAGHETFDIHR 445
Qy 423 AMFALKEKFEFTYINENF FAMQGPFAWRKSGIOMALGKRLHRYITQIVVYVQS 479
Db 445 ALNVAKAMPILPFWNININQILAAAPGSPKSGSLGELVYVQAFKMLQVNPNTID 502

RESULT 11
US-09-134-001C-4451
; Sequence 4451, Application US/09144001C
; Patent No. 6380370
; GENERAL INFORMATION:

```

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIORITY FILING DATE: 1998-08-13
PRIORITY FILING DATE: 1997-11-08
PRIORITY FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4451

LENGTH: 518

TYPE: PR1

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4451

Query Match 25.6%; Score 625.5; DB 4; Length 518;

Best Local Similarity 32.6%; Pred. No. 5 Se-56;

Matches 155; Conservative 98; Mismatches 198; Indels 25; Gaps 11;

8 PMVIGQFVTRGDAWIDVWPA-TEAVISIPDQQAEDAKAIDAAERAPPEALPAI 66

42 PLVINGEKLIT-KIDTF-NSVPANTISLIARVSKATQDIEKAFESANHAYQSWKKWSH 99

67 PRASWIRKISAGIRPASEISALIVEEGGKIQQLAEVEVAFADYDY-NAEMWARY 122

100 DPAEIIIPVAATIPRKEFISALIMVEAGKPMTEAGVDAEAGIDIEVYAPSMELA- 156

123 EGEIISDRPGENILIFKALGVITGILPWNFPFELIAPKMAPALLIGNIIVIKPSFT 182

157 DGRPV-LDPFGSEHNPYFKPIGTGVITPPNPPFPAIMAGTITLPAVAGNTVLLKPAUV 215

183 NNAIAFAKIVDEIGLPPRVFNILVGRGETVGTGFLAGNPKVAMVSMTGSVAGEKIMATAA 242

216 LIAYKIMEIFRAGIPGVVNFVPSDFKFGYVTHKGTHTVFTTGSFATGIVYLPISA 275

243 -----KNTKVKCLELGGKAPAIWMDADLELAVKAIIVDSRVINSOVNCARVYVQK 296

276 VVGRKQKFIKRVIAEMGGKDAIVDNNVLTDLAAEAVTSAPFSGGKSAKAIIVH 335

297 IYDQFVNLGSAWAVGQVGNIAERNDIAMGFLINAAALFKEVQKAVKAVEGAKVAG 356

336 VDEILEKALQITKLTIGNTEE--NTPMGVIVNKKOFUKIKNTIEGKREG-KLEFGG 392

357 AVEGKGYYPPTLLDVRQMSIMHEETFGVLEWVAFDITLEDIAISMANDSGLTSSII 416

393 TDSTGYPTIEPTIESGLQSAOKIMOEIIFGFWVGFIVKUDFKAIFVANDIDVGLTCAVI 452

417 TGNLAVAKAIKGLKEGTIYNRENFEAMQTFH--AGWPKSGISGAGSKHGLHLYL 470

453 TNHPRHWIKAVNEFDVGNILYNPGCTAAVVGYPFGFGKWS---TTCARTGSPVIL 506

RESULT 12

US-09-651-941-9

Sequence 9, Application US/09651941

Patent No. 6355470

GENERAL INFORMATION:

APPLICANT: ROUVIER, PIERRE F

APPLICANT: WALTERS, DANA M

APPLICANT: RAINER, RUSS

TITLE OF INVENTION: Genes Encoding Picric Acid Degradation

FILE REFERENCE: BC1022 US NA

CURRENT APPLICATION NUMBER: US/09/651,941

CURRENT FILING DATE: 2000-08-31

PRIORITY FILING DATE: 1997-11-08

PRIORITY FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 28

SOFTWARE: Microsoft Office 97

SEQ ID NO 9

LENGTH: 485

TYPE: PR1

ORGANISM: Rhodococcus erythropolis HL PM-1

US-09-955-597-9

Query Match 25.5%; Score 623.5; DB 4; Length 485;

Best Local Similarity 29.9%; Pred. No. 8e-56;

Matches 142; Conservative 99; Mismatches 225; Indels 9; Gaps 5;

8 PMVIGQFVTRGDAWIDVWPA-TEAVISIPDQQAEDAKAIDAAERAPPEALPAI 67

9 PLVIGDQITPSTGATFDSINPADGSHLASVAEATAADAVAFAPAAKAAAPTQMPHPAQ 68

68 KASWLKISAGIRPASEISALIVEEGGK-IQQQLAEVEVAFADYDYNAEMWARY 126

64 KIRIMFHYAALIFRHKITFLAQUSKUMKPIRESIGLIDLPIMILETFYFAGLVIRGR- 127

ORGANISM: Rhodococcus erythropolis HL PM-1

US-09-955-597-9

Query Match 25.5%; Score 623.5; DB 4; Length 485;

Best Local Similarity 29.9%; Pred. No. 8e-56;

Matches 142; Conservative 99; Mismatches 225; Indels 9; Gaps 5;

8 PMVIGQFVTRGDAWIDVWPA-TEAVISIPDQQAEDAKAIDAAERAPPEALPAI 67

9 PLVIGDQITPSTGATFDSINPADGSHLASVAEATAADAVAFAPAAKAAAPTQMPHPAQ 68

68 KASWLKISAGIRPASEISALIVEEGGK-IQQQLAEVEVAFADYDYNAEMWARY 126

64 KIRIMFHYAALIFRHKITFLAQUSKUMKPIRESIGLIDLPIMILETFYFAGLVIRGR- 127

127 TQSDRPGENI-LLFKRALGVITGILPWNFPFELIAPKMAPALLIGNIIVIKPSFTTNA 185

128 -TTPAPGRFLANTITLPEFGVVGATIPNMTAVQAVWKIAPALANGNAIVIKIAPLPLVP 186

186 IAFKIVDEIGLPPGVFNILVGRGETVGTGFLAGNPKVAMVSMTGSVAGEKIMATAAKNI 245

187 VALGELALEAGLPPULVNVLPGRGSAVAGNALVQHPESVQKVTFTUSTIEVQULGKMAADRL 246

246 TKVCELEGGKAPAIWMDADLELAVKAIIVDSRVINSOVNCARVYVQKIVYDQFVNRL 305

247 ITASLELGGKSAIAPGDSPPKAAVAVVPOAMYSNOGETTCTAPSRLIVERPIVDEVVHIV 306

306 GPMKAVGFGFNIAPRNDIAMGFLINAAALFKEVQKAVAFVAFVAFVAFVAFVAFVAFV 361

307 QAKVFAARVQVGLIP-PIHFTIGFISAPQKTSVRSYVVSSEIFRAILISGHSSEPPAPQ 365

362 GYYPPTILLDVRQMSIMHEETFGVLEWVAFDITLEDIAISMANDSGLTSSIIYTONLN 421

366 RYYPPTLFSGVTAIMPJAPETIFGPVIVLPPFEGEEELITLANDTVEGLAASVFTPEVG 425

422 VAMKAIKGLKEGTIYNRENFEAMQTFH--AGWPKSGISGAGSKHGLHLYL 476

424 PALRPAQTIDAGNVWINSWGVINPAGIYPPGSGGSGDLCQAAIESEFTREKSIW 480

RESULT 13

US-09-955-597-9

Sequence 9, Application US/09955597

Patent No. 6461856

GENERAL INFORMATION:

APPLICANT: ROUVIER, PIERRE E

APPLICANT: WALTERS, DANA M

APPLICANT: RAINER, RUSS

TITLE OF INVENTION: Genes Encoding Picric Acid Degradation

FILE REFERENCE: BC1022 US NA

CURRENT APPLICATION NUMBER: US/09/955,597

CURRENT FILING DATE: 2000-08-17

PRIORITY FILING DATE: 1997-11-08

PRIORITY FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 28

SOFTWARE: Microsoft Office 97

SEQ ID NO 9

LENGTH: 485

TYPE: PR1

ORGANISM: Rhodococcus erythropolis HL PM-1

US-09-955-597-9

Query Match 25.5%; Score 623.5; DB 4; Length 485;

Best Local Similarity 29.9%; Pred. No. 8e-56;

Matches 142; Conservative 99; Mismatches 225; Indels 9; Gaps 5;

8 PMVIGQFVTRGDAWIDVWPA-TEAVISIPDQQAEDAKAIDAAERAPPEALPAI 67

9 PLVIGDQITPSTGATFDSINPADGSHLASVAEATAADAVAFAPAAKAAAPTQMPHPAQ 68

68 KASWLKISAGIRPASEISALIVEEGGK-IQQQLAEVEVAFADYDYNAEMWARY 126

64 KIRIMFHYAALIFRHKITFLAQUSKUMKPIRESIGLIDLPIMILETFYFAGLVIRGR- 127

QY 127 IQSDRPGENT-LLFKRALGVTTGILFWNPFPFLIARKMAPALLTGNTIVIKPSEFTTNA 185
 DB 128 -TTPAPGRFLNYTLRPIGVVGAITFWNFPVAVQVWVKIAPALAMGNAVLKPAQALPLVP 186
 QY 186 IAFKIVDEIGLPRGVNVLGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNI 245
 DB 187 VALFETALFAGLPPGLVNVTPRGSVAAGNALVQHSVSKVTFNSELFWAGLQPMKAAQPL 246
 QY 246 TKVTEIGGRAPALVMLQALIELAVKAIIVSPVINSQW-N-AHVVYVUKGILYQFVNP 305
 DB 247 ITASLEIGKRSALVAFGDSKPAVAAVFPQAMYSNOGETCTAPSRLLVERPIYDEVVELV 306
 QY 306 GEAMQAVQFQNPAPERNTIANGPLINAAALFRVFKQVAPAVEEGARVFG----KAVEGK 361
 DB 307 QARVEARVGDPLD-PDTEIGPLISAEQRESVHSYVVSCTEGATLISGGDSQSPGAGEQ 365
 QY 362 GYYPPTLLIDVPQMSIMHEETEGVLPVAVFDLEDAISMANSDYGLTSSYTONIN 421
 DB 366 GFYRPTLFSGVADMRARIPERTFGPVLVLPFEGEEPAITLANIVFGLAAGVFTPDVG 425
 QY 422 VAMKAIKLFEGETYINRNFAMQ-FHAGTWKSKSLGGZALGKHJLGLYQTVVY 476
 DB 426 RALRFAOTLDAGNWNINSWCVLNPASPYRPGQSGYDGLGQAATESFTKEKSIW 480

RESULT 14
 US-09-655-270A-9
 ; Sequence 9, Application US/09655270A
 ; Patent No. 6329151
 ; GENERAL INFORMATION:
 ; APPLICANT: Rouviere, Pierre E.
 ; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic m
 ; FILE REFERENCE: H01011 US NA
 ; CURRENT APPLICATION NUMBER: US/09/655,270A
 ; CURRENT FILING DATE: 2000-09-05
 ; PRIOR APPLICATION NUMBER: 60/120,702
 ; PRIOR FILING DATE: 1999-February-19
 ; PRIOR APPLICATION NUMBER: 60/152,542
 ; PRIOR FILING DATE: 1999-September-03
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 9
 ; LENGTH: 508
 ; TYPE: PRT
 ; ORGANISM: Rhodococcus erythropolis HL PM-1
 US-09-655-270A-9

Query Match 35.5%; Score 623.5; DB 4; Length 508.
 Best local Similarity 29.9%; Pred. No. 8.6e-56;
 Matches 142; Conservative 99; Mismatches 225; Indels 9; Gaps 5;

QY 8 PMYIDGQFVTWRGDAMIDVWNPATEAVISKIPUGAELIARKAIUAARAPUEWALPAIE 67
 DB 32 PLVIGDGLTSSGATEDFSINPADGSHLASVAETAADUAVARAAKAAARTWQMRFAQ 91
 QY 68 RASWLKLSAGITRERASELSALIVBEGK-IQGLAEVEVAFATYIDYMAEWARYEGEI 126
 DB 92 RTRLMFRYALTEEHKTELAQOSRDMDCKPIRESGLDPLIMETLFYFAGLVTKIEGR- 150
 QY 127 IQSDRPGENT-LLFKRALGVTTGILFWNPFPFLIARKMAPALLTGNTIVIKPSEFTTNA 185
 DB 151 -TTPAPGRFLNYTLRPIGVVGAITFWNFPVAVQVWVKIAPALAMGNAVLKPAQALPLVP 209
 QY 186 IAFKIVDEIGLPRGVNVLGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNI 245
 DB 210 VALGELALEAGLPPGLVNVLPGRGSGVAGNALVQHSVSKVTFNSELFWAGLQPMKAAQPL 269
 QY 246 TKVTEIGGRAPALVMDADLELAVKAIIVSDSVINSQVCNCAERKVVYVUKGILYQFVNP 305
 DB 270 ITASLEIGKRSALVAFGDSKPAVAAVFPQAMYSNOGETCTAPSRLLVERPIYDEVVELV 329
 QY 306 GEAMQAVQFQNPAPERNTIANGPLINAAALFRVFKQVAPAVEEGARVFG----KAVEGK 361

DB 330 QARVEARVGDPLD-PDTEIGPLISAEQRESVHSYVVSCTEGATLISGGDSQSPGAGEQ 388
 QY 362 GYYPPTLLIDVPQMSIMHEETEGVLPVAVFDLEDAISMANSDYGLTSSYTONIN 421
 DB 389 GFYRPTLFSGVADMRARIPERTFGPVLVLPFEGEEPAITLANIVFGLAAGVFTPDVG 448
 QY 422 VAMKAIKLFEGETYINRNFAMQ-FHAGTWKSKSLGGZALGKHJLGLYQTVVY 476
 DB 449 RALRFAOTLDAGNWNINSWCVLNPASPYRPGQSGYDGLGQAATESFTKEKSIW 480

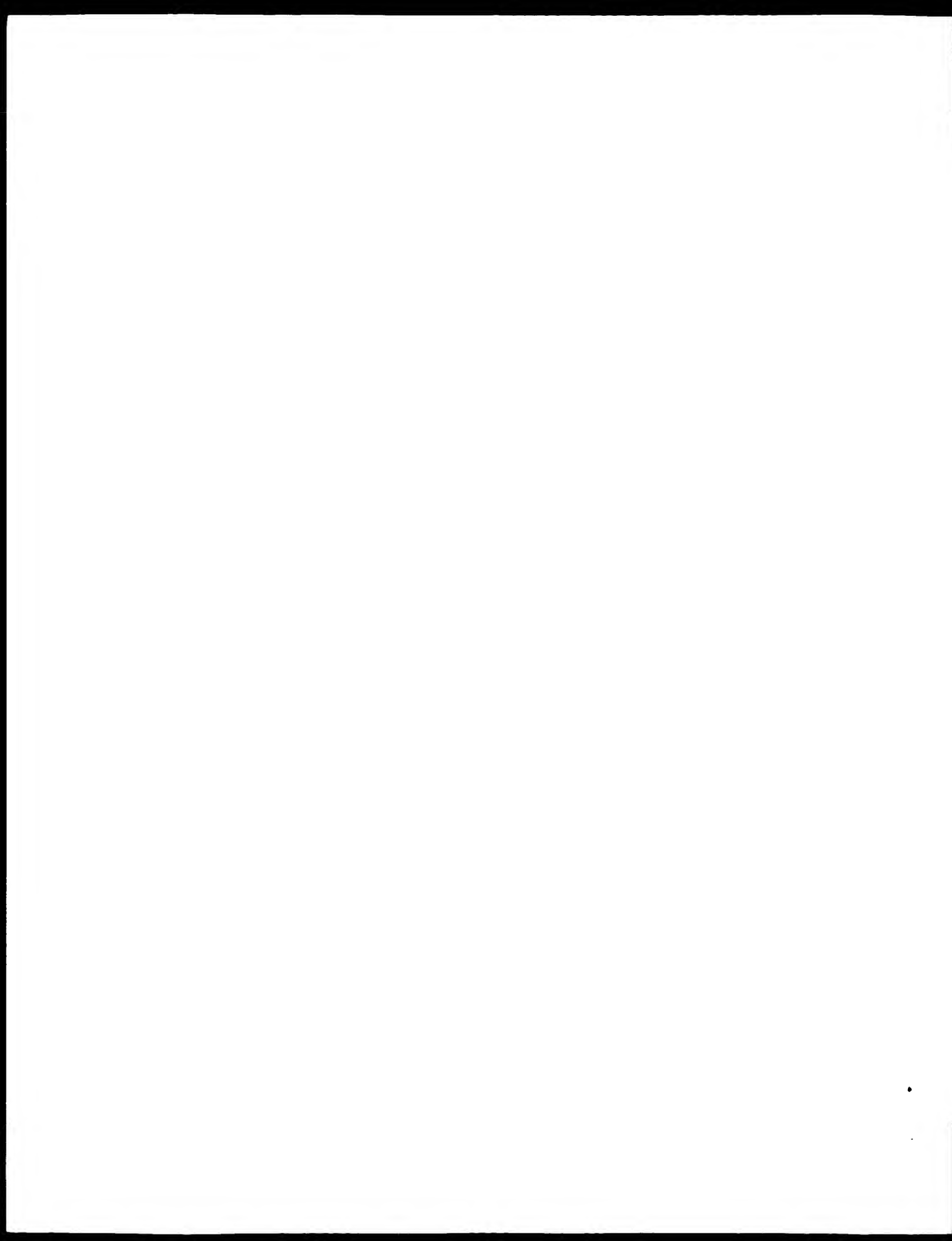
RESULT 15
 US-09-351-224E-5
 ; Sequence 5, Application US/09351224E
 ; Patent No. 6388171
 ; GENERAL INFORMATION:
 ; APPLICANT: Dwyck, Jon
 ; APPLICANT: Maddox, Joyce
 ; APPLICANT: Gilliam, Jacob
 ; APPLICANT: Folkerts, Otto
 ; APPLICANT: Grastia, Oswald K.
 ; TITLE OF INVENTION: Compositions and Methods for Eumonisin
 ; FILE REFERENCE: Detoxification
 ; CURRENT APPLICATION NUMBER: US/09/351,224E
 ; CURRENT FILING DATE: 1999-07-12
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 487
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-351-224E-5

Query Match 18.3%; Score 448; DB 4; Length 487.
 Best local Similarity 27.9%; Pred. No. 1.3e-47;
 Matches 135; Conservative 85; Mismatches 212; Indels 12; Gaps 14;

QY 9 MYIDGQFVTWRGDAMIDVWNPATEAVISKIPUGAELIARKAIUAARAPUEWALPA 65
 DB 12 LFINNEFVSSKSGERLILINPWLESIVADIVIVANAAUDSVAASVUAKKGLWKKETG 71
 QY 66 ITHASWLKLSAGITRERASELSALIVBEGK-IQGLAEVEVAFATYIDYMAEWARYEGEI 124
 DB 72 AQPAACMLKFAADLAEFAENAEKLALESITGPRVSMITHEDIPNMVSVFERYVAMAKIAG 131
 QY 125 EITQSDRPGENTILFKRALGVTTGILFWNPFPFLIARKMAPALLTGNTIVIKPSEFTTNN 184
 DB 132 KTEPELNGCKINWRY-EPMEGVVAGIASWNAIFLYVWKLALALAAQSFTRASKESKSLG 189
 QY 186 ALAPAKIVDEIGLPRGVNVLGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNI 243
 DB 190 VLGELALEAGLPPGLVNVLPGRGSGVAGNALVQHSVSKVTFNSELFWAGLQPMKAAQPL 248
 QY 244 NITKVTEIGGRAPALVMDADLELAVKAIIVSDSVINSQVCNCAERKVVYVUKGILYQFV 302
 DB 249 NMKRVTLLEDEK-PTIVNEAFLEPQSGESAPD-FSPPTLFWVHGS-LLVWV 309
 QY 303 NRGAMQAVQFQ-----NVARNDIANGPLINAAALFRVFKQVAPAVEEGARV 361
 DB 300 -NLAEPKPHVCHGSGQDGLVQGNLQPKR-THRPVVKSGYDGLVNTI-DVCKITQAL 356
 QY 352 AFGKRAVEKGYVYPPPTLLIDVWQKMSIMHEETEGVLPVAVFDLEDAISMANSDYGL 411
 DB 357 LTGVGKKGKQKPALEPTIFVNPKPGSKLWFEETFGPVLSTKTERFEREALIGANDIYGL 416
 QY 412 TSSYTONINVAAMKAIKLFEGETYINRNFAMQ-FHAGTWKSKSLGGZALGKHJLGLYQ 471
 DB 417 ASVITKSLNGPLGPPVSSALETSVSNITFFITFTCTPFGPKQSGSGPELDEEELKATLE 476
 QY 472 TQVV 475

Dk 477 PKT1 480

Search completed: June 24, 2003, 10:31:42
Job time : 15.2654 secs



GenCore version 5.1.6
Copyright (c) 1994 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 : Search time 23.1541 seconds
(without alignments)
2238 528 Million cell updates/sec

Title: US-09-830-751-6
Perfect score: 2446
Sequence: 1 MSVPVQHPMTISQFVTKGALSKHSLHAYLGLGVVYLS 479

Scoring table: KUSNM62
Gapop 10.0, Gapext 0.5

Searched 41772 seqs, 16826813 residues 41779
Total number of hits satisfying chosen parameters

Minimum DB seq length: 6
Maximum DB seq length: 200000000

Post-processing: Minimum Match 8%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_AA:*
- 1: /cgn2_6/protdata/2/pubpaa/us08_NEW_PUB pep.*
 - 2: /cgn2_6/protdata/2/pubpaa/PCT_NEW_PUB pep.*
 - 3: /cgn2_6/protdata/2/pubpaa/us06_NEW_PUB pep.*
 - 4: /cgn2_6/protdata/2/pubpaa/us06_PUB pep.*
 - 5: /cgn2_6/protdata/2/pubpaa/us07_NEW_PUB pep.*
 - 6: /cgn2_6/protdata/2/pubpaa/us07_PUB pep.*
 - 7: /cgn2_6/protdata/2/pubpaa/PCTUS_PUB pep.*
 - 8: /cgn2_6/protdata/2/pubpaa/us08_PUB pep.*
 - 9: /cgn2_6/protdata/2/pubpaa/us09_NEW_PUB pep.*
 - 10: /cgn2_6/protdata/2/pubpaa/us09_PUB pep.*
 - 11: /cgn2_6/protdata/2/pubpaa/us10_NEW_PUB pep.*
 - 12: /cgn2_6/protdata/2/pubpaa/us10_PUB pep.*
 - 13: /cgn2_6/protdata/2/pubpaa/us60_NEW_PUB pep.*
 - 14: /cgn2_6/protdata/2/pubpaa/us60_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Query Match	Score	Length	DB	ID	Description
1	879.5	36.0	483	10	US-09-815-242-5058	Sequence 5058, Ap
2	873.5	35.7	482	10	US-09-815-242-10264	Sequence 10264, A
3	858.5	35.1	482	10	US-09-815-242-14047	Sequence 14047, A
4	830	33.9	492	9	US-10-268-518-4	Sequence 4, Appl
5	819.5	33.5	493	9	US-10-175-696-21	Sequence 21, Appl
6	819.5	33.5	493	10	US-09-823-901-9	Sequence 9, Appl
7	771.5	31.5	496	10	US-09-815-242-5644	Sequence 5644, Ap
8	771.5	31.5	496	10	US-09-815-242-12657	Sequence 12657, A
9	768.5	31.4	481	10	US-09-815-242-13829	Sequence 13829, A
10	762	31.2	496	10	US-09-815-242-19550	Sequence 10550, A
11	756.5	30.9	475	10	US-09-815-242-12375	Sequence 12375, A
12	748.5	30.5	501	9	US-09-444-882-20	Sequence 20, Appl
13	748.5	30.5	501	9	US-10-293-865-22	Sequence 22, Appl
14	729	29.8	490	9	US-09-738-626-6572	Sequence 6572, Ap
15	729	29.8	490	10	US-09-815-242-12102	Sequence 12102, A
16	719.5	29.4	512	9	US-09-461-403-12	Sequence 12, Appl
17	719.5	29.4	512	9	US-10-268-518-2	Sequence 2, Appl
18	719.5	29.4	512	9	US-10-205-823-14	Sequence 14, Appl
19	712	29.1	490	10	US-09-815-242-10057	Sequence 10057, A

20	712	29.1	508	1	US-08-781-986A-5241	Sequence 5241, Ap
21	710	29.0	518	9	US-09-419-039-143	Sequence 143, Appl
22	706.5	28.9	534	9	US-09-444-882-24	Sequence 24, Appl
23	706.5	28.9	534	9	US-10-293-865-24	Sequence 24, Appl
24	700.5	28.6	538	9	US-09-444-882-20	Sequence 20, Appl
25	700.5	28.6	538	9	US-10-293-865-20	Sequence 20, Appl
26	691.5	28.3	487	9	US-10-175-696-17	Sequence 17, Appl
27	691.5	28.3	487	10	US-09-823-901-5	Sequence 5, Appl
28	681	27.8	496	9	US-09-847-208-59	Sequence 59, Appl
29	670	27.4	500	9	US-10-166-087-4	Sequence 4, Appl
30	662.5	27.1	498	9	US-10-272-419-12	Sequence 12, Appl
31	657.5	26.9	495	9	US-09-847-208-11	Sequence 11, Appl
32	643.5	26.3	491	10	US-09-997-664-3	Sequence 3, Appl
33	643.5	26.3	491	10	US-09-733-583-4	Sequence 4, Appl
34	636	26.0	504	9	US-09-738-626-6572	Sequence 6572, Ap
35	625	25.6	484	9	US-09-748-625-6430	Sequence 6430, Ap
36	610	24.9	454	9	US-09-748-626-4037	Sequence 4037, Ap
37	597	24.4	481	1	US-08-978-303-28	Sequence 28, Appl
38	597.5	24.3	434	10	US-09-815-232-13516	Sequence 13516, A
39	578.5	23.7	407	9	US-09-738-626-4108	Sequence 4108, Ap
40	529	21.5	521	9	US-09-748-625-6482	Sequence 6482, Ap
41	498	20.4	314	10	US-09-925-300-1348	Sequence 1348, Ap
42	484.5	19.6	504	9	US-09-748-625-6580	Sequence 6580, Ap
43	457.5	18.7	412	10	US-09-925-302-736	Sequence 736, Appl
44	448	18.3	487	9	US-09-882-694-5	Sequence 5, Appl
45	445.5	18.2	492	10	US-09-741-694-405	Sequence 406, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-5058
Sequence 5058, Application US/09/815/242
Patent NO. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Irawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815/242
PRIOR FILING DATE: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/243,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/259,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PastSeq for Windows Version 4.0
SEQ ID NO 5058
LENGTH: 483
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US 09 815 242-5058
Quality Match 36.04, Score 879.5, DB 10, Length 483.
Best Local Similarity 29.4, Pctd N 2 9.68;

[illegible]

```
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-3644

Query Match      31.5%, Score 771.5; DB 10; Length 496;
Best Local Similarity 36.4%, Pred. No. 6.2e-59;
Matches 170; Conservative 90; Mismatches 200; Indels 7; Gaps 3;

QY 10 YIDGQFVWPGDAMIDVNVNFAEAVISKIPGSAEAKAKAIDAAEPA--QPEWEALPAIE 67
DB 12 YIDGWEVSANKNRDRIINYNGEIVFIVSEGTAEAEKAILAARPAFESGMSQETAET 71
QY 68 RASMLGPKISAPIRKASRISALIVEEGRKIQGLAEVAFATAYITVMAEWAQKPYBSEII 127
DB 72 RGKKVPALADKIKHRRALAPLETITDGTGKTLSESYADMDIHNVMVYFACLADKDGEMI 131
QY 128 QSDRPGENIILFKPALGVTTGILPNWPPFFLIAPKMAPALLTGTNTIVIKPSEFTNNATA 187
DB 132 DSPIDPTESKIVKEPVGVVTTQITPNYPLLOQASWKIAPALATGCSLVMKPSRITPLTTIR 191
QY 188 FAKIVDEIGLPRGVFNVLGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNITK 247
DB 192 VFELMEEVGFPKGTINILLAGSEVGDVMSGHKEVDLVSTGQETGKILMKNAANNVIN 251
QY 248 VCELGGRKAPALVMDADLELAVKAIIVDSRVNSGVQNCACAEVYVQKGIYDOFVNRIGE 307
DB 252 IALELGGKPNPIIFDDAGFELAVDQALNGGYFHAGVWCASGRILVUNSLKDKFEQALID 311
QY 308 AMQAVQGNPAERNDIANGPLINAAALPERVEQKVARAVEGSRVAFSGRAVE----GKGY 363
DB 312 RVKTKIKLNGFTA-DTEMGPVISTEHPNKIESYMDVAKAPASATIAVSKRPDRDDLDKDL 370
QY 364 YYPPTLLDVRQEMSIMHEETFGVLPVPAFDLEDAISMANSDDYGLTSSIVTONINVA 423
DB 371 FPEPVTITNCDTSKRIVEEVGFPVTVGEFTEQEAQLQANSLVGLAGAVFSKDLGKA 430
QY 424 MKAIKGLKFGETYINRENFEAMQGFHAGWRKSGIGGADGKHGLHGYL 470
DB 431 ORVANKKLGTWINDRHPYPAQAPWGGYKQSGIGRGLRGLEEVYL 477

RESULT 8
US-09-815-242-12657
; Sequence 12657, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: FLITPA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/263,308
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12657
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12657

Query Match      31.5%, Score 771.5; DB 10; Length 496;
Best Local Similarity 36.4%, Pred. No. 6.2e-59;
Matches 170; Conservative 90; Mismatches 200; Indels 7; Gaps 3;

QY 10 YIDGQFVWPGDAMIDVNVNFAEAVISKIPGSAEAKAKAIDAAEPA--QPEWEALPAIE 67
DB 12 YIDGWEVSANKNRDRIINYNGEIVFIVSEGTAEAEKAILAARPAFESGMSQETAET 71
QY 68 RASMLGPKISAPIRKASRISALIVEEGRKIQGLAEVAFATAYITVMAEWAQKPYBSEII 127
DB 72 RGKKVPALADKIKHRRALAPLETITDGTGKTLSESYADMDIHNVMVYFACLADKDGEMI 131
QY 128 QSDRPGENIILFKPALGVTTGILPNWPPFFLIAPKMAPALLTGTNTIVIKPSEFTNNATA 187
DB 132 DSPIDPTESKIVKEPVGVVTTQITPNYPLLOQASWKIAPALATGCSLVMKPSRITPLTTIR 191
QY 188 FAKIVDEIGLPRGVFNVLGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNITK 247
DB 192 VFELMEEVGFPKGTINILLAGSEVGDVMSGHKEVDLVSTGQETGKILMKNAANNVIN 251
QY 248 VCELGGRKAPALVMDADLELAVKAIIVDSRVNSGVQNCACAEVYVQKGIYDOFVNRIGE 307
DB 252 IALELGGKPNPIIFDDAGFELAVDQALNGGYFHAGVWCASGRILVUNSLKDKFEQALID 311
QY 308 AMQAVQGNPAERNDIANGPLINAAALPERVEQKVARAVEGSRVAFSGRAVE----GKGY 363
DB 312 RVKTKIKLNGFTA-DTEMGPVISTEHPNKIESYMDVAKAPASATIAVSKRPDRDDLDKDL 370
QY 364 YYPPTLLDVRQEMSIMHEETFGVLPVPAFDLEDAISMANSDDYGLTSSIVTONINVA 423
DB 371 FPEPVTITNCDTSKRIVEEVGFPVTVGEFTEQEAQLQANSLVGLAGAVFSKDLGKA 430
QY 424 MKAIKGLKFGETYINRENFEAMQGFHAGWRKSGIGGADGKHGLHGYL 470
DB 431 ORVANKKLGTWINDRHPYPAQAPWGGYKQSGIGRGLRGLEEVYL 477

RESULT 9
US-09-815-242-13829
; Sequence 13829, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: FLITPA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/263,308
; PRIOR FILING DATE: 2001-02-16
```

```

: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13829
: LENGTH: 481
: TYPE: PRT
: ORGANISM: Salmonella typhi
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(481)
: OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815,242-1,3829

```

[illegible]

RESULT 10
US/09-815-242 10550
Sequence 10550, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.01A
CURRENT APPLICATION NUMBER: US/09/815,
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21

```

? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-24
? PRIOR APPLICATION NUMBER: 60/207,127
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-24
? PRIOR APPLICATION NUMBER: 60/258,626
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,941
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,408
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14710
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO: 10550
? LENGTH: 496
? TYPE: PRT
? ORG: 09-815-249-10550
? ORGANISM: Enterococcus faecalis

```

Query Match	41.2%	Score	76.2	IR	10	Length	496
Best Local Similarity	35.0%	Prod. No.	4.2e	58			
Matches	168	Conservative	98	Mismatches	206	Indels	8
QY	3	VIVQHPMYIDCGQVTHRCIDAWLIVVRIATCAVISHIRIQVLEAFKALIAALAKRAQHPENLA	6.2	1	1	1	1
DB	10	VASSYQLYTNGEIVLGGSKMKIASYNINPKLAEFVIALNALVGVAVAGAFALIDKRL	6.9	1	1	1	1
QY	63	LPIATERASWLKPSAGTRKERASELSALIVEEGR	12.1	1	1	1	1
DB	70	VIVAVTPSNILFKALHTEFNGFHLAMVETLNGKPKKPKTQSILVIALASALHREYFAVLEK	12.9	1	1	1	1
QY	122	VEGELIQSDRCENILFKKALGVTTGILLHWNTPFLLIAPKMAQALLTGNTIVLPSEEL	18.1	1	1	1	1
DB	140	REGSVKHEKDTULSIV-KETFGVGVQVLTITWNEFLHKGAWKLAPALAGNTVVTHFSSST	18.8	1	1	1	1
QY	182	TNNALAFAKIVIELGLPPEVFNVLGSGEIVFQELACNFKVAMVSMTFSSVSARER	24.1	1	1	1	1
DB	189	SLSLELLEFKHQQV-LPGGVNLTITGCSGSSNGMYMLAHPEFKLAITSTFWGIVYVAAVA	24.7	1	1	1	1
QY	242	AKNTKVCYLEGGRKAPALVMDADLELAVKAIVDSKVINSGOVNCPAHEVVVYKRGITYQF	30.1	1	1	1	1
DB	248	ADPTLPATIFPRGSANITPFDANWKAIRGVGLGTEFRQGVQVAGSVVFWQSGITYQF	30.7	1	1	1	1
QY	352	VNELEAMAVAFQGNPAFNPALMCPILINAAALEPVPEQVAPAVETPAPAVAPKAVE	35.9	1	1	1	1
DB	358	VEALKKEIQAVNGQFHWKQVEMWCAQVNIHQLEELKYVEIGAKVQALLTIGQRELEN	36.6	1	1	1	1
QY	450	--KGYYPPTLLLVQFERSIMBELTGGLVVVAFHLELALSMANISQGLSSITYI	41.7	1	1	1	1
DB	457	GLDQKGLAPLHANGINIMQVAGEELGVAVLKEEFEEFVILKANSISGLGAGVAG	42.6	1	1	1	1
QY	418	QNLNVAMKATKGLKEGTYGNRENFAMQGFHGMWERSIGKAGKGLHRYQLQVQVVL	47.7	1	1	1	1
DB	437	QNLNVAVAGVETPQMMVNTVQGLPVALHGGYKFSGLQPELQKSMLELQKLVNTI	48.6	1	1	1	1

```

RESULT 11
US-09-815-242-12475
: Sequence 12475, Application US/09/015242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsson, Karl L.
: APPLICANT: Zylsied, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA-011A

```

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/141,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/266,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12375
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12375

```

```

Query Match 30.98; Score 756.5; DR 10; Length 475;
Best Local Similarity 37.23; Pred. No. 1.2e-57;
Matches 175; Conservative 90; Mismatches 194; Indels 11; Gaps 7;

QY 10 YIDGQFTWRGDWIDVWNPATBAVISPIDGQAEADARKAIDAAERPAQEWALPAIERA 69
Db 8 YINGEWESNETIEVINPATEBIVGKAKGNKADVDRAVEAADNVYLEFRTSVKREQ 67

QY 70 SWLPKISAGIPERASEISALIVEGGKIQOLAE-VEVAFTADYIDYMAEWAPRYESEIIQ 128
Db 68 ALLQKIVKEYENKGGIVQAITDELGAPLSLSEVHYQGLNHFVAARKALLNYEFE--- 124

QY 129 SDRPGENILFLKRALGVTTGILPWNFPFFLIARKMAPALLTNTIVIKPSEPTNNIAF 188
Db 125 -ERRGDD-LVYKEAIGVSGGLITPWNFPNTJLSKLAFAAGSPVVLKPSSETPFAAVIL 182

QY 190 AKIVFETGIPGVENIVLGPETVQGEIACNPKVAMVMTGVSAGCFKIMATAAKNTTKV 248
Db 183 AEIFKVGKGVGVENIVNGDAGVGNPLSEHPKVRMMSTGSGTSGSKIMEKAADPKKV 242

QY 249 CIRLGGKAPATVMDADNLELAVKAVICSPVINSQVNCNCAERVYVQKGIYDQFVNPLGFA 308
Db 243 SLELGKSPYIVLDDVDVKEAAKATTCVKVYNTGQVCTAGTRVLPENKIKDAFLAELEQ 302

QY 309 MCAYQGNPAFNPDIAMGPINAAALFVFEQKVARAVFEKAPVAFGSKA-VEG--KGYV 365
Db 303 FSOVEVGNPRF-FRTQVSPITSKKQFTQGVNINKGIEETAFILFYKSPGKPRGLEKGYFA 361

QY 366 PPTLLDVRQEMSTHMETFGPVLPPVAFDTLEDAISMANDSDYGLTSSITYONLVAMK 425
Db 362 PPTIFIVNDQMTIAQEIFGPMVSVITYNDLDEATQIANDTKYGLAGYVIGKQKETHK 421

QY 426 AIKGLKRGETYINRNFEMAGPFAAGWPKSNTIGAGAKKHLHGYLETQVYV 475
Db 422 VAE-SLEAGHVLINAGKPKDLFF QYKQSGLGKENGQYGLIEFFLEK6SI 470

```

```

RESULT 12
US-09-344-882-22
; Sequence 22, Application US/09/444882
; Patent No. US20030162137A
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry I

```

```

; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

```

```

Query Match 30.63; Score 748.5; DR 9; Length 501;
Best Local Similarity 34.58; Pred. No. 6.3e-57;
Matches 165; Conservative 95; Mismatches 203; Indels 15; Gaps 6;

QY 9 MYIDGQFTWRGDWIDVWNPATBAVISPIDGQAEADARKAIDAAERA--QPEWALPAI 66
Db 22 LFINGQFUDASCKTFEIIOPNGSEVIAIIEGQKEUULAVNAAAPYAFDHQDPWPMTCF 81

QY 67 ERASWLKISAGIPERASEISALIVEGGKIQOLAE-VEVAFTADYIDYMAEWARYEGE 125
Db 82 ERAKLINKFADLIEENIEELAKLDVAVGGKLFQGLKYADIPATAGHFRYNAGAAADKHGE 141

QY 126 IIOSDRPGENILFLKRALGVTTGILPWNFPFFLIARKMAPALLTNTIVIKPSEPTNNIA 185
Db 142 TLKMTQSLCYTLKPEIGVGNLIIPWNEFSIMFATKVPAPAMAAGCTVWKPAEQTSLSA 201

QY 186 IAFKIVDEITGLPRGVNVLVLRGETVGOELACNPKVAMVMTGVSAGCFKIM-ATAAKN 244
Db 202 LFYAHLSKKEAN:PNGVIN:VIGHSTAGAAATASHMIW:KVSTQSTQSTQSTQSTQSTQ 261

QY 245 ITKVCLELGGKAPATVMDADNLELAVKAVICSPVINSQVNCNCAERVYVQKGIYDQFVNP 304
Db 262 LKAVSEHLEKSPILLINADHFAAKALIGTPYNPREFVWASFPVFPVFPVFPVFPVFP 321

QY 305 LGAMQAVQGNPAFNPDIAMGPINAAALFVFEQKVARAVFEKAPVAFGSKAVGKGY 364
Db 422 LVFKAKMTWVQDPD-SIAPGSPVLPVKPFEKILSVIEHKNEGAIIIGGKAIQDKGYF 380

QY 365 YPPTLLDVRQEMSTHMETFGPVLPPVAFDTLEDAISMANDSDYGLTSSITYONLVAM 424
Db 381 IQPTIEADVTFDKTYQDETFPGVMSLMKFKTVFPGIKCANNTKYGLAAGHLSQDIDLIN 440

QY 425 KAIKGLKRGETYINRNFEMAGPFAAGWPKSNTIGAGAKKHLHGYLETQVYV 477
Db 441 TVSPSIRKAGIITWVN-----CYFGFDLDCPYGYKMKSGNCFESOMALINYLQTKSVVM 493

```

```

PFSULT 13
US 15 203 865-22
; Sequence 22, Application US/10293865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry I
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and

```

1 TITLE OF INVENTION: Acetyl CoA Levels in Plants
2 FILE REFERENCE: 217113
3 CURRENT APPLICATION NUMBER: US/10/293,865
4 PRIOR FILING DATE: 2002-11-13
5 PRIOR APPLICATION NUMBER: US 09/344,882
6 PRIOR FILING DATE: 1999-06-25
7 PRIOR APPLICATION NUMBER: US 60/050,717
8 PRIOR FILING DATE: 1998-06-26
9 NUMBER OF SEQ ID NOS: 38
10 SOFTWARE: PatentIn Ver. 3.1
11 SEQ ID NO 22
12 LENGTH: 501
13 TYPE: PR1
14 ORGANISM: Arabidopsis Thaliana
15 US-10-293-865-22

Query Match 30.6%; Score 748.5; DB 9; Length 501;
Best Local Similarity 34.5%; Pred. No. 6, 3e-57;
Matches 165; Conservative 95; Mismatches 203; Indels 15; Gaps 6.
QY 9 MYIDGQVTVWRGDWIDVNPATEAVISIPDQQAEDARKAIDAERA--QPEWEALPAI 66
DB 22 LFIGQFIDAAASKTETIDPNEGVIATIAHQKEDVYAVNAAPYAFHSGPWPPTGTF 81
QY 67 ERASWLPKISAGIRFPASISALIVEEGGKIQQLAE-VEVAFTADYIDYMAEAPRPYEGE 125
DB 82 ERAKLNKPADLFEENIEELAKLDVADGKLPQLGKYADIPATAGHFRYNAGAADKIHGE 141
QY 126 IIOSDRPGENLLFKKALGVYTGILPWNPFFLIARKMAPALLTNTIVKPSFTNNNA 185
DB 142 TLKMTKOSLPGYTLKEPIGPGWGNIIIPWNPFSIMFATKVPAPAMAAGCINWVKPAPOTSLSA 201
QY 186 IAFKIVDTEIGLPGGVNVLVGRGETVGGELAGNEKPVAMVSMTGSVSAKEKIM-ATAAKN 244
DB 202 LFVAHLSKEAGIPDGVNLNVTGFGSTAGAAIAASHMDVDKVSFTGSTDVGRKIMAAASRN 261
QY 245 ITKVCFLGLGKAPAVIMDDADLELAVKAIVDSRVINSQVNCABRVYVQKGYIDQFVNR 304
DB 262 LKKSLELGKSPLLIFNDADIDKAADLALLGCFYKNGELICVASSKVFVQGVYDKVVEK 321
QY 305 LGAMQAVQFNPABNDIAMGPLNMAALERVEQKVARAVEEGARVAFGCKAVGKGY 364
DB 322 LVKAKIWTVGDFD-STARQSPQVUKKQFELSYIEHKNEGATILTGKALIGDKGYF 380
QY 365 YPPTLLDVQKESIMHEETFGVLPVAVFDLEDAISMANDSDYGLTSSIVYTONLNVAM 424
DB 381 IOPTIEADVTEGMKIYODEIFGVVSLMKFKTVEEGIKCANNTKYGLAAGILSQDIDLIN 440
QY 425 KAIKGLKPGHTYINRENPFAMQCF-----HAGWRKSGIGGADGKHLGHTYLOTQVVYL 477
DB 441 TVSRIRAGILWVN-----CYFGFDIDCPYGGYKWSGNCRESGMDALDNLQTKSVVM 493

RESULT 14
US-09-738-626-3558
1 Sequence 3558, Application US/09738626
2 Publication No. US20020197605A1
3 GENERAL INFORMATION:
4 APPLICANT: NAKAGAWA, SATOSHI
5 APPLICANT: MIZOGUCHI, HIROSHI
6 APPLICANT: ANDO, SEIKO
7 APPLICANT: HAYASHI, MIKIO
8 APPLICANT: OCHIAI, KEIHO
9 APPLICANT: YOKOI, HARUHIKO
10 APPLICANT: TATEISHI, NAKO
11 APPLICANT: SENOH, AKIHRO
12 APPLICANT: IKEDA, MASATO
13 APPLICANT: OZAKI, AKIO
14 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
15 FILE REFERENCE: 249-125
16 CURRENT APPLICATION NUMBER: US/09/738,626
17 CURRENT FILING DATE: 2000-12-18
18 PRIOR APPLICATION NUMBER: JP 99/377484

1 PRIOR FILING DATE: 1999-12-16
2 PRIOR APPLICATION NUMBER: JP 00/159162
3 PRIOR FILING DATE: 2000-04-07
4 PRIOR APPLICATION NUMBER: JP 00/280988
5 PRIOR FILING DATE: 2000-08-03
6 NUMBER OF SEQ ID NOS: 7059
7 SOFTWARE: PatentIn ver. 3.0
8 SEQ ID NO 3558
9 LENGTH: 490
10 TYPE: PR1
11 ORGANISM: Corynebacterium glutamicum
12 US-09-738-626-3558

Query Match 29.8%; Score 729; DB 9; Length 490;
Best Local Similarity 34.4%; Pred. No. 4, 1e-55;
Matches 160; Conservative 89; Mismatches 206; Indels 10; Gaps 4.
QY 20 GDAWI-----DVNPATEAVISIPDQQAEDARKAIDAERAQPEWEALPAERASW 71
DB 19 GDSWVEASDGGTFDVENPATGETIATLASATSEIALAALDAACAVJALWAHMAPAKERSN 78
QY 72 LPKISWCLPEPASEISALIVEEGGKIQQLAFVFAFTADYIDYMAEAPRPYEGELIJSOR 141
DB 79 LRKGFELVAERAREEATILMEKMKELARAGEVYNGNEFLWFSSEAVRIGRYGHTPE 148
QY 132 PCENILLPKALGVYTGILPWNPFFLIARKMAPALLTNTIVKPSFTNNNAIATAKI 191
DB 139 GNLRLMTALKPVQPCILLITPWNPFLAMATRKVAVPALAAGVWVVKPAPLITTSQYFAC 198
QY 192 VIEETGPGGVNVLVGGELGTVGDELAVNPKVAMVSMTGSVSAKEKIMATAAKRITKVT 250
DB 199 MLIAGLPAPVNLVNSASASATSNPIMEDURKVSFDSIPVQGLTKKAADKAVIRISM 258
QY 251 FICGKAPAVIMDDADLELAVKAIVDSRVINSQVNCABRVYVQKGYIDQFVNRKIDAMQ 310
DB 259 ELCCNAPFIVFEDADIDIAIEGACMKKRNIDGAA-TAARITLVHSVADESRERFAKLE 318
QY 311 AVGEPNPAERNITAMPD-INAAALFVQKVARAVEEGARVAFGCKAVGKGYIDQFVNR 370
DB 319 EGVLGNCIDRG-VTVCPLVEERKARSDVASIVAAVAKCATVLITGCKATGACGYEVEPTVL 377
QY 371 IIVYQPMSTIMHETFGVLPVAVFDLEDAISMANDSDYGLTSSIVYTONLNVAMKAIKGL 400
DB 378 TGVSTUADILNEIEFGVAPVITFQTEEPALRLANSTEGVGLASVYFQTSKIRFVSIGL 437
QY 431 KGEIYINKENE-EAMQGF-HAMKKSQIGAGCKKRLHRYLOTQVV 474
DB 438 EPCLVVNSVVISNAAPFNVKVSQSGMRGEGELGIEYISQVYI 492

RESULT 15
US-09-815-242-12102
1 Sequence 12102, Application US/090415242
2 Patent No. US20020061569A1
3 GENERAL INFORMATION:
4 APPLICANT: Haselbeck, Robert
5 APPLICANT: Ohlsen, Kari L.
6 APPLICANT: Zyskind, Judith W.
7 APPLICANT: Wall, Daniel
8 APPLICANT: Trawick, John D.
9 APPLICANT: Carr, Grant J.
10 APPLICANT: Yamamoto, Robert T.
11 APPLICANT: Xu, H. Howard
12 TITLE OF INVENTION: Identification of Essential Genes in
13 TITLE OF INVENTION: Prokaryotes
14 FILE REFERENCE: FLITRA-011A
15 CURRENT APPLICATION NUMBER: US/09/815,242
16 CURRENT FILING DATE: 2001-03-21
17 PRIOR APPLICATION NUMBER: 60/191,078
18 PRIOR FILING DATE: 2000-03-21
19 PRIOR APPLICATION NUMBER: 60/206,848
20 PRIOR FILING DATE: 2000-05-24
21 PRIOR APPLICATION NUMBER: 60/207,722

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 12102
LENGTH: 490
TYPE: PRI
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12102

Query Match 29.8%, Score 729, DB 10, Length 490;
Best Local Similarity 34.2%; Pred. No. 3, le-55;
Matches 164, Conservative 101, Mismatches 206, Indels 8, Gaps 5;

QY 6 QHPYIDGQVYTWGDMIDVNPATPAVISRIIDGQEDARKAIDAERAQPEWALPA 65
DB 6 EQKLYIGGRVEASSGAFELINPANGVELAKVAKSKEDVEKAVQSAVEGQKVMAMIA 65
QY 66 IHHASWIKKISAGIHHKASEISALIVEE3K-IQGLAEVAVATADYIDYMAEMARREY 124
DB 66 MORSPTIIPPAVDILPEPNDELALETLDTSKPLAETPSVDITVGADVLEYYAGLVPAIEG 125
QY 125 EIIOSDRPENILFEKRALGVTTGILPWNPFELIARKMAPALLTNTIYIKPSEPTNN 184
DB 126 EQIPL-RETSFYTRREPLGVAGIGAMNYPVOIALMKSPALAGNAMIFKPEVTPLT 184
QY 185 AIAFAKIVDELGLRGVFNVLIGRGETVGOELAGNPKVAVSMTGSVSGEKIMATA-AK 243
DB 185 ALKLAELIYTAGVDPGVFNLTGSGREVGOMLTHEPLIEKISTGTGISTGKKVMASASS 244
QY 244 NITVVCLELGGKAPAIYMDADLELAVKAIYDSRVINSQOVNCAERVYQKGIYDQFVN 303
DB 245 SLKEVTMLGSKSPILIFPDADLDRKADIVMANFSSGQVCTNGTIRVFTHRSQARFEA 304
QY 304 RIIGAMQAVCPGNPAFRNDIAMSPTINAAI FVEQKVAFAVEGARVAFGKAVE---- 359
DB 305 KLEEVORIRIGDPQDEN-TNPGPLVSFPMESVLYTESGKAQKAKLLOGGERVUDGAF 363
QY 360 GKGYIYPTLLDLVQKESIMHEETGSPVLPVVAFDLLEDAISMANDSDYGLTSSITYON 419
DB 364 GKGAIVAPPTVTPCPDMITVFPRIFGPVMSIIVYDDEDAIRANDTEYGLAGVYTD 423
QY 420 INVAMKAIKGLKKESEIYNFENFEAMGPHACWPKSGTSGAKGKHSJASVLTGVVYIQ 479
DB 424 LARAHRAIHRLEAGICINNTMGESPALMPYGGYKQSGVGENGLITLAHTRIKSYOVE 482

Search completed: June 24, 2003, 10:37:06
Job time: 24.1541 secs

Genome version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model:

Run on: Tue Jun 24, 2003, 10:11:00 : Search time 17 x 414 seconds

(without alignments)
1780.862 Million cell updates/sec

Title: us-09-830-751-6

Perfect score: 2446
1 MSVPVQHPMYIDQCEVTWPGDAWIIIVVNATPFAVISPPIPGGAPFAPKAIIDAAERAPFW 60

Scoring table: BLASTSUM62

Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Pending Patents-AA-Main*

1 /cgn2_6/ptodata1/paa/us086_COMB.pep.*
2 /cgn2_6/ptodata1/paa/us087_COMB.pep.*
3 /cgn2_6/ptodata1/paa/us07_COMB.pep.*
4 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
5 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
6 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
7 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
8 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
9 /cgn2_6/ptodata1/paa/us04_COMB.pep.*
10 /cgn2_6/ptodata1/paa/us086_COMB.pep.*
11 /cgn2_6/ptodata1/paa/us087_COMB.pep.*
12 /cgn2_6/ptodata1/paa/us088_COMB.pep.*
13 /cgn2_6/ptodata1/paa/us089_COMB.pep.*
14 /cgn2_6/ptodata1/paa/us090_COMB.pep.*
15 /cgn2_6/ptodata1/paa/us091_COMB.pep.*
16 /cgn2_6/ptodata1/paa/us092_COMB.pep.*
17 /cgn2_6/ptodata1/paa/us093_COMB.pep.*
18 /cgn2_6/ptodata1/paa/us094_COMB.pep.*
19 /cgn2_6/ptodata1/paa/us095_COMB.pep.*
20 /cgn2_6/ptodata1/paa/us096_COMB.pep.*
21 /cgn2_6/ptodata1/paa/us097_COMB.pep.*
22 /cgn2_6/ptodata1/paa/us098_COMB.pep.*
23 /cgn2_6/ptodata1/paa/us099_COMB.pep.*
24 /cgn2_6/ptodata1/paa/us100_COMB.pep.*
25 /cgn2_6/ptodata1/paa/us101_COMB.pep.*
26 /cgn2_6/ptodata1/paa/us102_COMB.pep.*
27 /cgn2_6/ptodata1/paa/us103_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2432	99.4	21	US-09-791-537-112887
2	2432	99.4	22	US-09-830-751-6
3	2432	99.4	27	US-09-830-751-6
4	2157	88.7	18	US-09-489-039A-13820
5	1557	63.7	21	US-09-791-537-10455
6	1552	63.5	21	US-09-791-537-115281

Result No.	Score	Query Match Length	ID	Description
7	1499	61.3	16	US-09-251-631-5824
8	1499	61.3	16	US-09-251-631-5824
9	1279.5	52.3	19	US-09-791-537-20873
10	942.5	48.1	18	US-09-450-069-4782
11	941.5	38.1	14	US-09-730-279-2956
12	903	40.3	27	US-09-460-039-13571
13	900	36.8	27	US-09-460-039-13571
14	894	36.5	21	US-09-791-537-115281
15	879.5	36.0	22	US-09-815-242-10264
16	879.5	36.0	22	US-09-815-242-10264
17	879.5	36.0	22	US-09-815-242-10264
18	879.5	36.0	22	US-09-815-242-10264
19	879.5	36.0	22	US-09-815-242-10264
20	879.5	36.0	22	US-09-815-242-10264
21	879.5	36.0	22	US-09-815-242-10264
22	879.5	36.0	22	US-09-815-242-10264
23	879.5	36.0	22	US-09-815-242-10264
24	879.5	36.0	22	US-09-815-242-10264
25	879.5	36.0	22	US-09-815-242-10264
26	879.5	36.0	22	US-09-815-242-10264
27	879.5	36.0	22	US-09-815-242-10264
28	879.5	36.0	22	US-09-815-242-10264
29	879.5	36.0	22	US-09-815-242-10264
30	879.5	36.0	22	US-09-815-242-10264
31	879.5	36.0	22	US-09-815-242-10264
32	879.5	36.0	22	US-09-815-242-10264
33	879.5	36.0	22	US-09-815-242-10264
34	879.5	36.0	22	US-09-815-242-10264
35	879.5	36.0	22	US-09-815-242-10264
36	879.5	36.0	22	US-09-815-242-10264
37	879.5	36.0	22	US-09-815-242-10264
38	879.5	36.0	22	US-09-815-242-10264
39	879.5	36.0	22	US-09-815-242-10264
40	879.5	36.0	22	US-09-815-242-10264
41	879.5	36.0	22	US-09-815-242-10264
42	879.5	36.0	22	US-09-815-242-10264
43	879.5	36.0	22	US-09-815-242-10264
44	879.5	36.0	22	US-09-815-242-10264
45	879.5	36.0	22	US-09-815-242-10264

ALIGNMENTS

RESULT 1

US-09-791-537-112887

Sequence: 112887, Applicant: Escherichia coli

GENERAL INFORMATION:

APPLICANT: Biocompare, Inc.

APPLICANT: Debe, Derek

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

FILE REFERENCE: 251/210

CURRENT APPLICATION NUMBER: US/09/791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ. IDS: 153055

SOFTWARE: PatentIn version 3.0

SEQ ID NO 112887

LENGTH: 479

TYPE: PRT

ORGANISM: Escherichia coli

US-09-791-537-112887

Query Match: 99.4%, Score: 2432, DR: 21, Length: 479.

Best Local Similarity: 99.6%, Pred. No.: 8.4e-233.

Matches: 477; Conservative: 0; Mismatches: 2; Indels: 0; Gaps: 0.

Query: 1 MSVPVQHPMYIDQCEVTWPGDAWIIIVVNATPFAVISPPIPGGAPFAPKAIIDAAERAPFW 60

DB: 1 MSVPVQHPMYIDQCEVTWPGDAWIIIVVNATPFAVISPPIPGGAPFAPKAIIDAAERAPFW 60

61 EATPAPFASWLPKISGATIRFASPLSALIVFQKQIQGAPVAVFAVPIIVVMATWAP 120

```

|||||
Db      61  EALPAIERASMLRKISAGIIPERASEISALIVEEGCKTQOLAEVFAFTADYIDYAEAR 120
Qy      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Db      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Qy      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Db      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Qy      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Db      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Qy      401  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 460
Db      401  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 460
Qy      461  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Db      461  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Qy      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479
Db      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479

RESULT 2
US-09-801-563-9
: Sequence 9, Application US/09801563
: GENERAL INFORMATION:
: APPLICANT: Levy, Stuart, et. al.
: TITLE OF INVENTION: NIMH COMPOSITIONS AND THEIR METHODS OF USE
: FILE REFERENCE: PKZ-043
: CURRENT FILING DATE: 2001-03-08
: PRIOR APPLICATION NUMBER: 60/188,362
: PRIOR FILING DATE: 2000-04-10
: NUMBER OF SEQ ID NOS: 98
: SOFTWARE: Paton In. Ver. 2.0
: SEQ ID NO 9
: LENGTH: 479
: TYPE: PRT
: ORGANISM: Escherichia coli
US-09-801-563-9

Query Match 99.4%; Score 2432; DB 27; Length 479;
Best Local Similarity 99.6%; Pred. No. 8.4e-233;
Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  MSYVQHPMTIDQGFVTWRGDAMIDVYNATFAVISRIPDQGAEDARAKIDAAERADPEW 60
Db      1  MSYVQHPMTIDQGFVTWRGDAMIDVYNATFAVISRIPDQGAEDARAKIDAAERADPEW 60
Qy      61  EALPAIERASMLRKISAGIIPERASEISALIVEEGCKTQOLAEVFAFTADYIDYAEAR 120
Db      61  EALPAIERASMLRKISAGIIPERASEISALIVEEGCKTQOLAEVFAFTADYIDYAEAR 120
Qy      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Db      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Qy      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Db      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Qy      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Db      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Qy      401  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 460
Db      401  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 460
Qy      461  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Db      461  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Qy      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479
Db      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479

```

```

Db      301  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 360
Qy      361  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Db      361  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Qy      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479
Db      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479

RESULT 2
US-60-360-039-823
: Sequence 823, Application US/60360039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianfeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 36-10(52052)A
: CURRENT FILING DATE: US/60/360,039
: PRIOR APPLICATION NUMBER: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 823
: LENGTH: 479
: TYPE: PRT
: ORGANISM: Escherichia coli
US-60-360-039-823

Query Match 99.4%; Score 2432; DB 27; Length 479;
Best Local Similarity 99.6%; Pred. No. 8.4e-233;
Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  MSYVQHPMTIDQGFVTWRGDAMIDVYNATFAVISRIPDQGAEDARAKIDAAERADPEW 60
Db      1  MSYVQHPMTIDQGFVTWRGDAMIDVYNATFAVISRIPDQGAEDARAKIDAAERADPEW 60
Qy      61  EALPAIERASMLRKISAGIIPERASEISALIVEEGCKTQOLAEVFAFTADYIDYAEAR 120
Db      61  EALPAIERASMLRKISAGIIPERASEISALIVEEGCKTQOLAEVFAFTADYIDYAEAR 120
Qy      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Db      121  KYEEELIUSDRPGENILLEKRALGVTTGILPMNPFPLIARKMAPALLTGNTVIKPSF 180
Qy      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Db      181  TTNNALAFKIVETIGLPGVFNILVIGKEETVGGELAGNPKVAMVSMTGSVSAEKIMAT 240
Qy      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Db      241  AAKNITKVCLELGKAPALVMDADLELAVKALVDSRVINSQGVNCAERVVYQKGYDQ 300
Qy      301  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 360
Db      301  FVNRLGEAMQAVQGFNPAERNDIAMGPLINAAALERVQKVAARAVEGARAFAVGKAVG 360
Qy      361  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Db      361  KGYVPPPTLLDVROKMSIMHEETPGPVLPVVAFTLEDAISMANDSYGLTSSITYONL 420
Qy      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479
Db      421  NVAMKAIKGLKFGETYINRENFAMQGFHAGMKRSIGADCKHGLHETLOTQVYVYLOS 479

RESULT 4
US-09-489-039A-13820
: Sequence 13820, Application US/09489039A
: GENERAL INFORMATION:

```

APPLICANT: Gary Breton et. al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 FILE REFERENCE: PFM0001AE FOR DIAGNOSTICS AND THERAPEUTICS
 CURRENT APPLICATION NUMBER: 2703.2004001
 CURRENT FILING DATE: US/09/489, 039A
 PRIOR APPLICATION NUMBER: 2000-01-27
 PRIOR FILING DATE: US 60/117,747
 NUMBER OF SEQ ID NOS: 1999-01-29
 SEQ ID NO: 13820
 LENGTH: 514
 TYPE: PRT
 ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-13820

Query Match 88.2%; Score 2157; DB 18; Length 514;
 Best Local Similarity 85.8%; Pred. No. 2,66-205;
 Matches 411; Conservative 43; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSVPQHPMYIDGQFVTRGDMIDVNPATFATISIPGQADAKATDAERQPEW 60
 DB 36 MTAFOHPMTIDGQFVTRGDMIDVNPATFATISIPGQADAKATDAERQPEW 95
 QY 61 EALPAIERASMLRKISAGIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAP 120
 DB 96 EALPAIERASMLRKISAGIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAP 155
 QY 121 REEGTIDSDRGEHILLFRAICVTTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEF 180
 DB 156 REEGTIDSDRGEHILLFRAICVTTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEF 215
 QY 181 TTNNAIAFAKIVDEIGLPRGVFNILVGRGETVGGELGNPKVAVNSMTGSAGEKIMAT 240
 DB 216 TPNNAIARAVIWHVQGLKGVFNILVGRGETVGGELGNPKVAVNSMTGSAGEKIMAT 275
 QY 241 AAKNTTKVCLGSKAPAIYMDADLELAIVKAVIVESFVINSQVQNCAPFVYVQKSLYDQ 300
 DB 276 AAKNTTKVCLGSKAPAIYMDADLELAIVKAVIVESFVINSQVQNCAPFVYVQKSLYDQ 335
 QY 301 FVNRIQFAMQAVQGNPNPFRNDIAMPPLINAAALFVREGVAVAVFEGCAVAFGSKAVET 360
 DB 336 FVNRIQFAMQAVQGNPNPFRNDIAMPPLINAAALFVREGVAVAVFEGCAVAFGSKAVET 395
 QY 361 KGYVFPITLLDVRKEMSIMHEELFGEVLVPAVDLEDAISMANDSDYGLISYIQLN 420
 DB 396 KGYVFPITLLDVRKEMTIHETHPGYVLVPAVSTLDELAISMANDSDYGLISYIQLN 455
 QY 421 NVAMKATIGSKFGETYINPENEAMQGHAWKSKSIIQSAHKKHGIHGYLQTVVYLYGS 479
 DB 456 NVAMKATIGSKFGETYINPENEAMQGHAWKSKSIIQSAHKKHGIHGYLQTVVYLYGS 514

RESULT 5
 US-09-791-537-40455
 Sequence 40455, Application US/09791537
 GENERAL INFORMATION:
 APPLICANT: Biogenix, Inc.
 APPLICANT: Debe, Derek
 APPLICANT: Dancer, Joseph
 TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
 FILE REFERENCE: 261/210
 CURRENT APPLICATION NUMBER: US/09/791,537
 CURRENT FILING DATE: 2001-02-22
 NUMBER OF SEQ ID NOS: 153055
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO: 40455
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Neisseria meningitidis MC58
 US-09-791-537-40455

Query Match 63.7%; Score 1557; DB 21; Length 480;

Best Local Similarity 63.2%; Pred. No. 1,7e-145;
 Matches 301; Conservative 63; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHPMYIDGQFVTRGDMIDVNPATFATISIPGQADAKATDAERQPEW 65
 DB 3 QIAMIYINPHTENLEHMKPIVNLNSTRFALIAHKKKQKALVPAVAVAAAPAPAFMEFLIA 62
 QY 66 IERASMLRKISAGIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAREE 125
 DB 63 VEKGAYLKRIAGIIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAREE 122
 QY 126 ITOSRPGENTLLKRPALVGTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEPTNNA 185
 DB 123 ITOSRPGENTLLKRPALVGTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEPTNNA 182
 QY 186 IAPAKIVDEIGLPRGVFNILVGRGETVGGELGNPKVAVNSMTGSAGEKIMATAKNI 245
 DB 183 HIFAEIVDVGGLPAGVFRVNVGPGAEIGNALSHQGVDMVSLTGSVEACRGVMAASANI 242
 QY 246 TKVCLDEGSKAPAIYMDADLELAIVKAVIVESFVINSQVQNCAPFVYVQKSLYDQ 305
 DB 243 TKVCLDEGSKAPAIYMDADLELAIVKAVIVESFVINSQVQNCAPFVYVQKSLYDQ 302
 QY 306 GEAMQAVQGNPNPFRNDIAMPPLINAAALFVREGVAVAVFEGCAVAFGSKAVET 363
 DB 303 TAAMKQVYVNPAPAEHNALEEMPLEIEPAVKAIVAEVPAVAVQSAKLVQVHAKSKAV 362
 QY 364 YVPITLLDVRKEMSIMHEELFGEVLVPAVDLEDAISMANDSDYGLISYIQLN 423
 DB 363 YVPITLLDVRKEMTIHETHPGYVLVPAVSTLDELAISMANDSDYGLISYIQLN 422
 QY 424 MKALIKGLKSKFGETYINPENEAMQGHAWKSKSIIQSAHKKHGIHGYLQTVVYLYGS 479
 DB 423 FYVIRKIDGGETYINPENEAMQGHAWKSKSIIQSAHKKHGIHGYLQTVVYLYGS 478

RESULT 6
 US-09-791-537-115281
 Sequence 115281, Application US/09791537
 GENERAL INFORMATION:
 APPLICANT: Biogenix, Inc.
 APPLICANT: Debe, Derek
 APPLICANT: Dancer, Joseph
 TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 FILE REFERENCE: 261/210
 CURRENT APPLICATION NUMBER: US/09/791,537
 CURRENT FILING DATE: 2001-02-22
 NUMBER OF SEQ ID NOS: 153055
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO: 115281
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Neisseria meningitidis Z2491
 US-09-791-537-115281

Query Match 63.5%; Score 1552; DB 21; Length 480;
 Best Local Similarity 63.0%; Pred. No. 5.5e-145;
 Matches 300; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHPMYIDGQFVTRGDMIDVNPATFATISIPGQADAKATDAERQPEW 65
 DB 3 QIAMIYINPHTENLEHMKPIVNLNSTRFALIAHKKKQKALVPAVAVAAAPAPAFMEFLIA 62
 QY 66 IERASMLRKISAGIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAREE 125
 DB 63 VEKGAYLKRIAGIIPERASETSALIVEEGSKTQQLAEVAFATADYIYMAEMAREE 122
 QY 126 ITOSRPGENTLLKRPALVGTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEPTNNA 185
 DB 123 ITOSRPGENTLLKRPALVGTGIIIPMPFPFLIARKKAPALLTNTIIVIKSEPTNNA 182
 QY 186 IAPAKIVDEIGLPRGVFNILVGRGETVGGELGNPKVAVNSMTGSAGEKIMATAKNI 245

```

184 HIFAFIDVAGLAFAGVNVNNGAGLGNLSAHPOVDWVSLTGSSEARQYMEASANI 242
185
246 TKVTELEGGKAPATVMTATLELAVKAVIYSPVNSGVNNAFVYVYKGTTCGPFVNL 305
247
248 TKSLTELEGGKAPATVMTATLELAVKAVIYSPVNSGVNNAFVYVYKGTTCGPFVNL 302
249
406 GFMAOVQVGNP--AFERNIDJAMGDLINMAALEHVRQKVARAVEGKAVAPVYKKAFFGKY 363
407
408 TAAKAVYKGNVAKALGALGKPLLEKAVKAVAKVAVKAVKAGKLVGGRKRAFGKY 362
409
464 YVPTLLAVKQKMSIMHEETFGVLPVVAAPDLLEAISMANSYGLTSSITONLVA 423
465
466 FFEPTLLIDNSMDIMKEETFGVLPVVAAPDLLEAISMANSYGLTSSITONLVA 422
467
424 MKAKIKLKGKGTITINNEFEAMGCHACMKKSLGADKRGHGLYLTUVVYLS 479
425
426 FVYIRIOPGDETYINNEFEAMGCHACMKKSLGADKRGHGLYLTUVVYLS 478

```

RESULT 7

```

US-09-252-691-5824
: Sequence 5824, Application US/09252691B
: GENERAL INFORMATION:
: APPLICANT: Keith G. Weinstein et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
: FILE REFERENCE: 107196.135
: CURRENT APPLICATION NUMBER: US/09/252.691B
: CURRENT FILING DATE: 1999-02-18
: NUMBER OF SEQ ID NOS: 11324
: SEQ ID NO 5824
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Enterobacter cloacae
US-09-252-691-5824

```

Query Match

```

Best Local Similarity 61.3% Score 1499; DB 16; Length 319;
Matches 292; Conservative 17; Mismatches 10; Indels 0; Gaps 0;

```

```

161 RKMAPALTGNTIVIKPSEFTNNNAIAFAKIVDEIGLPGGVNVLVIGGFTVGOELAGNP 220
162
221 KVAWVSMTGSVSGEKITATKANKITKVLTLEGGKAPATVMTATLELAVKAVIYSPVNS 280
222
61 KVAWVSMTGSVSGEKITATKANKITKVLTLEGGKAPATVMTATLELAVKAVIYSPVNS 120
62
281 SGVVCNCAEYVYOKGIYKFNKRGKAAKAVQFNPFAERNDIAMGPLINAAALERVYOK 340
282
121 TGVVCNCAEYVYOKGIYKFNKRGKAAKAVQFNPFAERNDIAMGPLINAAALERVYOK 180
122
441 VARAVEGKAVAFGKAVGKGYVYPTLLIDVROKMSIMHEETFGVLPVVAAPDLLEA 400
442
181 VARAVEGKAVAFGKAVGKGYVYPTLLIDVROKMSIMHEETFGVLPVVAAPDLLEA 240
182
401 ISMANSYGLTSSITONLVAAMKAIKGLKFGTYINNEFEAMGCHACMKKSLGADK 460
402
241 LNMANSYGLTSSITONLVAAMKAIKGLKFGTYINNEFEAMGCHACMKKSLGADK 300
242
461 DCKHGLNGYLTUVVYLS 479
462
301 DCKHGLNGYLTUVVYLS 478

```

RESULT 8

```

US-09-252-691-5824
: Sequence 5824, Application US/09252691B
: GENERAL INFORMATION:
: APPLICANT: Keith G. Weinstein et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
: FILE REFERENCE: 107196.135
: CURRENT APPLICATION NUMBER: US/09/252.691B
: CURRENT FILING DATE: 1999-02-18
: NUMBER OF SEQ ID NOS: 11324
: SEQ ID NO 5824
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Enterobacter cloacae
US-09-252-691-5824

```

```

: FILE REFERENCE: 107196.135
: CURRENT APPLICATION NUMBER: US/09/252.691B
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/094,145
: PRIOR FILING DATE: 1998-07-24
: PRIOR APPLICATION NUMBER: US 60/074,787
: PRIOR FILING DATE: 1998-02-18
: NUMBER OF SEQ ID NOS: 11326
: SEQ ID NO 5824
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Enterobacter cloacae
US-09-252-691-5824

```

Query Match

```

Best Local Similarity 61.3% Score 1499; DB 16; Length 319;
Matches 292; Conservative 17; Mismatches 10; Indels 0; Gaps 0;

```

```

161 RKMAPALTGNTIVIKPSEFTNNNAIAFAKIVDEIGLPGGVNVLVIGGFTVGOELAGNP 220
162
221 KVAWVSMTGSVSGEKITATKANKITKVLTLEGGKAPATVMTATLELAVKAVIYSPVNS 280
222
61 KVAWVSMTGSVSGEKITATKANKITKVLTLEGGKAPATVMTATLELAVKAVIYSPVNS 120
62
281 SGVVCNCAEYVYOKGIYKFNKRGKAAKAVQFNPFAERNDIAMGPLINAAALERVYOK 340
282
121 TGVVCNCAEYVYOKGIYKFNKRGKAAKAVQFNPFAERNDIAMGPLINAAALERVYOK 180
122
441 VARAVEGKAVAFGKAVGKGYVYPTLLIDVROKMSIMHEETFGVLPVVAAPDLLEA 400
442
181 VARAVEGKAVAFGKAVGKGYVYPTLLIDVROKMSIMHEETFGVLPVVAAPDLLEA 240
182
401 ISMANSYGLTSSITONLVAAMKAIKGLKFGTYINNEFEAMGCHACMKKSLGADK 460
402
241 LNMANSYGLTSSITONLVAAMKAIKGLKFGTYINNEFEAMGCHACMKKSLGADK 300
242
461 DCKHGLNGYLTUVVYLS 479
462
301 DCKHGLNGYLTUVVYLS 478

```

RESULT 9

```

US-09-791-537-20873
: Sequence 20873, Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Debe, Derek
: APPLICANT: Danzer, Joseph
: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
: FILE REFERENCE: 261/210
: CURRENT APPLICATION NUMBER: US/09/791,537
: CURRENT FILING DATE: 2001-02-22
: NUMBER OF SEQ ID NOS: 153055
: SOFTWARE: Patent version 3.0
: SEQ ID NO 20873
: LENGTH: 393
: TYPE: PRT
: ORGANISM: Campylobacter jejuni
US-09-791-537-20873

```

Query Match

```

Best Local Similarity 52.3% Score 1279.5; DB 21; Length 393;
Matches 245; Conservative 52; Mismatches 90; Indels 1; Gaps 1;

```

```

31 VERGKTLQIAVEVAFAIYVYMAEWAAPVYGFILISHPSPENHILFKRAIGVTTGIL 150
32
1 MOEGKTRVLASTIEINFTADYMDYAEWARRYEGEIIOSDRANSHIYVLSAIGVIGIL 60
151 PMWFPEPLIAPPMAPATIGNTIVIKPSEFTNNNAIAFAKIVDEIGLPGGVNVLVIGG 210
152

```

```

Db      61  PKNPFPLIAPKMAPALLTGNCTVIKPSSEFPNNAFEKLVSSSLPKGVENLVACKCS 120
Qy      211 TVGGLACNPKVAMSGTGSVAGEKIMATAKNTVCLGSKAPAIYMDADDELA 270
      121 VVGYELSSNENIGWILSGVSEAGTVMEEAAKNIIVSELELGKAPAIYCKADIDLA 180
Qy      271 KAVISPVYNSGVNCFNERYVYQKJIDJVFYNLLSLAMLAJVCUNAFENULAMGLIN 330
      181 EAKISKRLGNNGVCKNAPRAVHTSVYEFVCKEYKAMSKSVSNLT-KADEMGSELVN 239
Qy      331 AAALREVECKARAVEGARAFAFGKAVEGKYVPTLIDVFOEISMHEETGEVLP 390
      240 GAGVINALAMLOKATAKAIYECOKKITGSGYFFHPSVLINVKHELEIMOKELPAILP 299
Qy      391 VVAFTLEALSMANDSYGLSSITVONLVAKAKIKGKCEYINKENFEAMOGFHA 450
      300 IAKFTLDEVIDMADCEYGLSSITVONLIDIAMRASHKIKPPTVINPENFEAMOGFHA 359
Qy      451 GMRKSGIGADGKHGHGYLOTGVVYLD 478
      360 GFRKSGIGADGKHGHGHLYLAHVYLD 347

```

```

RESULT 10
US-09-450-969-4782
Sequence 4782, Application US/09450969
GENERAL INFORMATION:
APPLICANT: Iyon Doucette-Stamm et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: PAT99-09A
CURRENT APPLICATION NUMBER: US/09/450,969
CURRENT FILING DATE: 1999-11-29
NUMBER OF SEQ ID NOS: 7544
SEQ ID NO 4782
LENGTH: 493
TYPE: PRT
ORGANISM: S.epidermidis
US-09-450-969-4782

```

```

Query Match      38.1%; Score 932.5; DB 18; Length 493;
Best Local Similarity 38.5%; Pred. No. 3.8e-83;
Matches 181; Conservative 108; Mismatches 180; Indels 1; Gaps 1;

```

```

Qy      9  MYIDGFYTPGDADIVVNPATREAVISPIFGGAPKRAITPAEFAQPEWEALPIFER 68
      24 LEINNEFLESQKEMVYNPATGAEPTTLTALIEEVNDAIEKSCQAOJEMERVPQTR 63
      69 ASWLPKISAGIREPASEISALIVBERCKTQGLAFVVAFTADYIIVMAEMARREYEGPT 128
      84 AEHVKLIPLEKRNDEIAQIVYKQCKTLAQAVGLIKSISFLDYMSLSMSKSPVLD 143
Qy      129 SRRPGPNITLFRKRAIGVTGTLIPMNPFFLIARMAFALLTGNITVYKPSSEFTNNNAIAF 188
      144 NSTIANETIQINKPLGIVAGIVPMNAPILVIMKRVIAIVGGSVVIKPSSEFTLLTLKL 203
      189 AKIVDEILPFGVENVLVRGETVGOELAGNPKVAMSGTGSVAGEKIMATAKNTKY 248
      204 AELPASTIPATGIFVIGTGTETVIGTGLASHKIDQILSLTSMKAVKSVENNAQIVKVV 263
Qy      249 LTLIVVQGENSMHEETGEVLPVVAFTLEALSMANDSYGLSSITVONLVAKAKIK 428
      264 NLELGGNAPVIVTSNADLKRAVNTVAPINNAGVNTCPERIVHEVHDDFLNKVISK 323
      309 MAAVQFQNPABEENDIAMGPIINAAALFEVFGKVAFAFEFARVAFEGKAVGKGYVYPT 368
      324 MSLIVGQRPDEN-TDYGALLINGKQSTHREKVDATKKNATIKTGGHOLKRHGFYAPT 342
Qy      369 LTLIVVQGENSMHEETGEVLPVVAFTLEALSMANDSYGLSSITVONLVAKAKIK 428
      383 VILNPKIVNVEKELEFEVLALITYPDEFVLEVLANDINGLSSYIFSENLFEVMTATE 442
Qy      429 GIKFGFTYINPENFAMOGFHAQWMSKSIKAKKHLHSHVYLDGVVYLD 478

```

```

Db      443 RLKFEVYANCAEEVANGYHAGMRBSGLGADGIGHFEETVYNTVSYTR 492

```

```

RESULT 11
US-09-710-279-2956
Sequence 2956, Application US/09710279
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P0348005
CURRENT APPLICATION NUMBER US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2956
LENGTH: 479
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-2956

```

```

Query Match      38.1%; Score 931.5; DB 21; Length 479;
Best Local Similarity 38.5%; Pred. No. 4.6e-83;
Matches 181; Conservative 109; Mismatches 179; Indels 1; Gaps 1;

```

```

Qy      9  MYIDGFYTPGDADIVVNPATREAVISPIFGGAPKRAITPAEFAQPEWEALPIFER 68
      10 LEINNEFLESQKEMVYNPATGAEPTTLTALIEEVNDAIEKSCQAOJEMERVPQTR 69
      69 ASWLPKISAGIREPASEISALIVBERCKTQGLAFVVAFTADYIIVMAEMARREYEGPT 128
      70 AEHVKLIPLEKRNDEIAQIVYKQCKTLAQAVGLIKSISFLDYMSLSMSKSPVLD 129
Qy      129 SRRPGPNITLFRKRAIGVTGTLIPMNPFFLIARMAFALLTGNITVYKPSSEFTNNNAIAF 188
      130 NSTIANETIQINKPLGIVAGIVPMNAPILVIMKRVIAIVGGSVVIKPSSEFTLLTLKL 189
      189 AKIVDEILPFGVENVLVRGETVGOELAGNPKVAMSGTGSVAGEKIMATAKNTKY 248
      190 AELPASTIPATGIFVIGTGTETVIGTGLASHKIDQILSLTSMKAVKSVENNAQIVKVV 249
Qy      249 LTLIVVQGENSMHEETGEVLPVVAFTLEALSMANDSYGLSSITVONLVAKAKIK 428
      250 NLELGGNAPVIVTSNADLKRAVNTVAPINNAGVNTCPERIVHEVHDDFLNKVISK 409
      309 MAAVQFQNPABEENDIAMGPIINAAALFEVFGKVAFAFEFARVAFEGKAVGKGYVYPT 368
      324 MSLIVGQRPDEN-TDYGALLINGKQSTHREKVDATKKNATIKTGGHOLKRHGFYAPT 348
Qy      369 LTLIVVQGENSMHEETGEVLPVVAFTLEALSMANDSYGLSSITVONLVAKAKIK 428
      429 RLKFEVYANCAEEVANGYHAGMRBSGLGADGIGHFEETVYNTVSYTR 478

```

```

RESULT 12
US-60-360-039-18199
Sequence 18199, Application US/60360039
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Chen, Xianfeng
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```

Query Match	46.6%; Score 900; DB 27; Length 473;
Best Local Similarity	48.5%; Pctd. No. 6, 2 nd 80,
Matches	182; Conservative 95; Mismatches 188; Indels 8; Gaps 2;

Query Match	36.5%;	Score 894;	DB 21;	Length 492;
Best Local Similarity	38.9%;	Pred. No. 2,66-79;		
Matches 184;	Conservative 92;	Mismatches 111;	Indels 6;	Gaps 3;
QY	10	YITGGVITKGLAMILEVYNPALEAVISITPLQAGLEAKKAIIDAAKQYBWEALALAIKKA	69	
DB	19	FVAGFEELDSSGTLPLVIPANSGDIAVEPSSSTIVPAVTAAGACGEMSRHSVAVA	78	
QY	70	SWIPIKTSATIPPEPASETSALIVPEPKIQGLAFEFVAFATVYIYMAEWAPYEETIOS	129	
DB	79	AVILEAVRODAIVAHAEIARISVFOGKTLISAKETHTGACAFDPAISQRYAAGSMMAS	138	
QY	130	DPEGENILLEFKALGVTTGILPMNPPPELLAKKMAPALLTGNTVIRKSEFTTNALIFA	189	
DB	139	E-PGRSLGVEEDPICVVAAILPMNFPVALIPARKVAPALMAGNAVVLAKPSELTLPLSALLA	197	
QY	190	KIYDELIGIPRGVFNIVIGRGFTVGGGLAGNKKVAMYSMTSVSAGKKIMATLAKKNITVGC	249	
DB	198	FLCPLAIVPPGILASVVTERRKDTIPALVTHIRVGMATMTISTRRGELLIAVAADQIIPVS	257	

1

2

GenCore version 5.1.6
Copyright (c) 1994 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:35, Search time 51.8555 seconds
(without alignments)
2375.712 Million cell updates/sec

Title: US-09-830-751-6

Sequence: 1 MSVPVGHMYITFGFVITWKA... ALNRKHLNRITDQIVVITUS 479

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1171708 seqs, 257189365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Pending Patents: AA-New: *
1: /cgn2_6/plodata/2/paa/ECU_NEW_COMB.pep.*
2: /cgn2_6/plodata/2/paa/BS06_NEW_COMB.pep.*
3: /cgn2_6/plodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/plodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/plodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/plodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/plodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2446	100.0	479	US-09-830-751-6	Sequence 63, Appl 1
2	2432	99.4	479	US-10-369-493-823	Sequence 83, App
3	2157	88.2	514	US-10-446-203-13820	Sequence 13820, A
4	1552	63.5	480	US-10-282-122A-65716	Sequence 65716, A
5	1499	61.3	319	US-10-417-HM6-5824	Sequence 5824, Ap
6	1279.5	52.3	393	US-10-282-122A-54340	Sequence 54340, Ap
7	932.5	38.1	493	US-10-092-411A-4388	Sequence 4388, Ap
8	903	36.9	493	US-10-369-493-18199	Sequence 18199, A
9	900.5	36.8	480	US-10-282-122A-65716	Sequence 65716, A
10	900.5	36.8	480	US-10-282-122A-65716	Sequence 65716, A
11	900	36.0	473	US-10-369-493-13957	Sequence 13957, A
12	879.5	36.0	483	US-10-282-122A-43455	Sequence 43455, A
13	879.5	36.0	594	US-10-365-683-27749	Sequence 27749, A
14	879.5	36.0	594	US-10-419-128-27749	Sequence 27749, A
15	873.5	35.7	482	US-10-282-122A-56650	Sequence 56650, A
16	873.5	35.7	482	US-10-369-493-23511	Sequence 23511, A
17	868.5	35.5	482	US-10-282-122A-75217	Sequence 75217, A
18	863.5	35.3	478	US-10-369-493-13957	Sequence 13957, A
19	863.5	35.3	480	US-10-282-122A-56650	Sequence 56650, A
20	859.5	35.1	465	US-10-369-493-7938	Sequence 7938, Ap
21	859.5	35.1	482	US-10-282-122A-75217	Sequence 75217, A
22	853.5	34.9	482	US-10-282-122A-59544	Sequence 59544, A
23	853.5	34.9	487	US-10-446-203-13820	Sequence 13820, A
24	845.5	34.6	475	US-10-369-493-17412	Sequence 17412, A
25	836	34.2	483	US-10-282-122A-46972	Sequence 46972, A
26	835	34.1	468	US-10-369-493-17164	Sequence 17164, A

27	833.5	34.1	489	US-10-282-122A-51224	Sequence 51224, A
28	842	44.0	484	US-10-282-122A-77428	Sequence 67328, A
29	831.5	34.0	485	US-10-369-493-11863	Sequence 11863, A
30	831	34.0	458	US-10-369-493-10934	Sequence 10934, A
31	830	33.9	492	US-10-282-122A-75217	Sequence 4, Appl 1
32	820	33.9	492	US-10-282-122A-75217	Sequence 4, Appl 1
33	820	33.9	492	US-10-282-122A-75217	Sequence 4, Appl 1
34	824	33.7	478	US-10-369-493-501	Sequence 501, App
35	822	33.6	482	US-10-369-493-17031	Sequence 17031, A
36	821.5	33.5	479	US-10-369-493-7523	Sequence 7523, Ap
37	821.5	33.5	481	US-10-369-493-4763	Sequence 4763, Ap
38	821.5	33.5	486	US-10-282-122A-49742	Sequence 49742, A
39	816	33.4	477	US-10-369-493-21219	Sequence 21219, A
40	815	33.3	537	US-10-424-599-162810	Sequence 162810, A
41	814.5	33.3	551	US-10-417-063-125211	Sequence 125211, A
42	813.5	33.3	482	US-10-369-493-8449	Sequence 8449, Ap
43	813.5	33.3	485	US-10-369-493-12251	Sequence 12251, A
44	813.5	33.3	489	US-10-282-122A-50658	Sequence 50658, A
45	812	33.2	460	US-10-369-493-16681	Sequence 16681, A

ALIGNMENTS

RESULT 1
US-09-830-751-6
Sequence 6, Application US/09830751
GENERAL INFORMATION:
APPLICANT: Suthers, Patrick F.
APPLICANT: Cameron, Douglas C.
TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant
FILE REFERENCE: 2002/06, 2001/7
CURRENT APPLICATION NUMBER: US/09-830,751
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/151,440
PRIOR FILING DATE: 1999-08-30
PRIOR APPLICATION NUMBER: PCT/US00/23878
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 479
TYPE: PRT
ORGANISM: Escherichia coli
US-09-830-751-6

Query Match 100.0%; Score 2446; DB 5; Length 479;
Best Local Similarity 100.0%; Pred. No. 1.9e 205;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Caps 0;

1 MSVPVGHMYITFGFVITWKA... ALNRKHLNRITDQIVVITUS 60
1 MSVPVGHMYITFGFVITWKA... ALNRKHLNRITDQIVVITUS 60
61 EALFATFASWLFKISACIFEFASFTSLIVETQYFGLAEVFAVTA-YITVMFAW 120
61 PALPATERASWLFKISACIFEFASFTSLIVETQYFGLAEVFAVTA-YITVMFAW 120
121 REGGEIISDRGENILIFKRALGVTTGILPMPNPFELIARKKAPALLTGTITIKPSEF 180
121 REGGEIISDRGENILIFKRALGVTTGILPMPNPFELIARKKAPALLTGTITIKPSEF 180
181 TRNNATAPATVDFGIFPGVNIYVLSGELVGOELANPNVAVMSMGVSNGFKTMAF 240
181 TRNNATAPATVDFGIFPGVNIYVLSGELVGOELANPNVAVMSMGVSNGFKTMAF 240
241 AAFNTIKVTLFELGKAPATVMTATCTAVNAVITSPVINSQVTRAFEVVYAKTITPD 300
241 AAKNIIKVLTLGSKAPATVMTATCTAVNAVITSPVINSQVTRAFEVVYAKTITPD 300
301 FVNLGEMQAVQFNPAFBNITAMGRIINAALEFVQKAPAVGCAVAVRGAVER 360
301 FVNLGEMQAVQFNPAFBNITAMGRIINAALEFVQKAPAVGCAVAVRGAVER 360

DB 401 FVNHLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 360
 QY 461 KGYVPTLLDVROKMSIMHEETFGVLPVVAFTLEDAISMANISQGLTSSITQNL 420
 DB 461 KGYVPTLLDVROKMSIMHEETFGVLPVVAFTLEDAISMANISQGLTSSITQNL 420
 QY 421 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 479
 DB 421 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 479

RESULT 2

US-10-469-493-823
 ? Sequence 823, Application US/10469493
 ? GENERAL INFORMATION:
 ? APPLICANT: Cao, Yompei
 ? APPLICANT: Hiroko, Gregory J.
 ? APPLICANT: Slater, Steven G.
 ? APPLICANT: Goldman, Barry S.
 ? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ? FILE REFERENCE: 48-10(52052)H
 ? CURRENT APPLICATION NUMBER: US/10469, 493
 ? PRIOR FILING DATE: 2003-02-28
 ? PRIOR APPLICATION NUMBER: US 60/460, 039
 ? PRIOR FILING DATE: 2002-02-21
 ? NUMBER OF SEQ ID NOS: 47374
 ? SEQ ID NO 823
 ? LENGTH: 479
 ? TYPE: PRT
 ? ORGANISM: Escherichia coli
 US-10-469-493-823

Query Match 99.4%; Score 2432; DB 6; Length 479;
 Best Local Similarity 99.6%; Pred. No. 3, 26-204;

Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVYVQHPMTIDQVYTWKGDAMIIVVNPATFAVISIRIDGQAEARAKIDAERAQDEW 60
 DB 1 MSVYVQHPMTIDQVYTWKGDAMIIVVNPATFAVISIRIDGQAEARAKIDAERAQDEW 60
 QY 61 EALPATERASWLRKISAGIRERASELSALIVEGCKIQDLAVEVAFTADYIDYMAEMAR 120
 DB 61 EALPATERASWLRKISAGIRERASELSALIVEGCKIQDLAVEVAFTADYIDYMAEMAR 120
 QY 121 RYEEETIUSDRPSENTILFKRAIGVTTGILPWNPFPELLARKMAPALLTGNTVIKPEEF 180
 DB 121 RYEEETIUSDRPSENTILFKRAIGVTTGILPWNPFPELLARKMAPALLTGNTVIKPEEF 180
 QY 181 TTNNALAPAKTYDEILPGGVFNLYVGRGETVGOELAGNPKYAMSMGTSVAGEKIMAT 240
 DB 181 TTNNALAPAKTYDEILPGGVFNLYVGRGETVGOELAGNPKYAMSMGTSVAGEKIMAT 240
 QY 241 AAKNITKVCLELGKAPATVMDADLELAVKATVDSRVINSQVNCFAERYVQGIYDQ 300
 DB 241 AAKNITKVCLELGKAPATVMDADLELAVKATVDSRVINSQVNCFAERYVQGIYDQ 300
 QY 401 FVNRLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 360
 DB 401 FVNRLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 360
 QY 461 KGYVPTLLDVROKMSIMHEETFGVLPVVAFTLEDAISMANISQGLTSSITQNL 420
 DB 461 KGYVPTLLDVROKMSIMHEETFGVLPVVAFTLEDAISMANISQGLTSSITQNL 420
 QY 421 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 479
 DB 421 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 479

RESULT 3
 US-10-446-203-14820

? Sequence 13820, Application US/10446203
 ? GENERAL INFORMATION:
 ? APPLICANT: Gary Breton et al
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ? FILE REFERENCE: 2709, 2004001
 ? CURRENT APPLICATION NUMBER: US/10446, 203
 ? PRIOR FILING DATE: 2003-05-27
 ? PRIOR APPLICATION NUMBER: US/09/489, 039
 ? PRIOR FILING DATE: 2000-01-27
 ? PRIOR APPLICATION NUMBER: US 6,011,747
 ? PRIOR FILING DATE: 1999-01-29
 ? NUMBER OF SEQ ID NOS: 14342
 ? SEQ ID NO 13820
 ? LENGTH: 514
 ? TYPE: PRT
 ? ORGANISM: Klebsiella pneumoniae
 US-10-446-203-13820

Query Match 88.2%; Score 2157; DB 6; Length 514;
 Best Local Similarity 85.8%; Pred. No. 4, 66-180;

Matches 411; Conservative 43; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSVYVQHPMTIDQVYTWKGDAMIIVVNPATFAVISIRIDGQAEARAKIDAERAQDEW 60
 DB 36 MTAPOVHPMTIDQVYTWKGDAMIIVVNPATFAVISIRIDGQAEARAKIDAERAQDEW 95
 QY 61 EALPATERASWLRKISAGIRERASELSALIVEGCKIQDLAVEVAFTADYIDYMAEMAR 120
 DB 46 EALPATERASWLRKISAGIRERASELSALIVEGCKIQDLAVEVAFTADYIDYMAEMAR 155
 QY 121 RYEEETIUSDRPSENTILFKRAIGVTTGILPWNPFPELLARKMAPALLTGNTVIKPEEF 180
 DB 156 RYEEETIUSDRPSENTILFKRAIGVTTGILPWNPFPELLARKMAPALLTGNTVIKPEEF 215
 QY 181 TTNNALAPAKTYDEILPGGVFNLYVGRGETVGOELAGNPKYAMSMGTSVAGEKIMAT 240
 DB 216 TTNNALAPAKTYDEILPGGVFNLYVGRGETVGOELAGNPKYAMSMGTSVAGEKIMAT 275
 QY 241 AAKNITKVCLELGKAPATVMDADLELAVKATVDSRVINSQVNCFAERYVQGIYDQ 300
 DB 276 AAKNITKVCLELGKAPATVMDADLELAVKATVDSRVINSQVNCFAERYVQGIYDQ 335
 QY 301 FVNRLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 360
 DB 336 FVNRLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 395
 QY 401 FVNRLEFAMQAVQFQGNIAERNDIAMGHTLINAALIFRQGVAAVEEFAVAFGKAVES 360
 DB 396 KGYVPTLLDVROKMSIMHEETFGVLPVVAFTLEDAISMANISQGLTSSITQNL 455
 QY 421 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 479
 DB 456 NVAMKAIKGIKRGETYINRENFAMQGFHAGKRSKISGADGKHGLHGYLOTUVVYLOS 514

RESULT 4

US-10-282-122A-65716
 ? Sequence 65716, Application US/10282122A

? GENERAL INFORMATION:
 ? APPLICANT: Wang, Jiansu
 ? APPLICANT: Zamudio, Carlos
 ? APPLICANT: Malone, Cheryl
 ? APPLICANT: Haselbeck, Robert
 ? APPLICANT: Ohlsen, Kari
 ? APPLICANT: Zyskind, Judith
 ? APPLICANT: Wall, Daniel
 ? APPLICANT: Trawick, John
 ? APPLICANT: Carr, Grant
 ? APPLICANT: Yamamoto, Robert
 ? APPLICANT: Forsyth, R.
 ? APPLICANT: Xu, H.
 ? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```

: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282.122A
: PRIORITY FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-25
: PRIOR APPLICATION NUMBER: 60/210,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,925
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ. ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ. ID NO: 65716
: LENGTH: 480
: TYPE: PRT
: ORGANISM: Neisseria meningitidis
US-10-282-122A-65716

```

```

Query Match 63.5% Score 1552; DB 6; Length 480;
Best Local Similarity 64.0% Pctd No. 4 3e-127;
Matches 300; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHPYIGQFVTMGDAIDVNPATEAVISIPDQAEFAKFAIDAEAPQEWELDA 65
DB 3 QLAYINGREFENDENGMRVNLMPSTEEAIAFRPKGKADVBAAVAAAPRAMERLPA 62
QY 66 IERASMLPKTISAGIREPASTISATVEFGKIQQLAEVEVAFTADYIDYMAEMARPEGE 125
DB 63 VERGAVYLRKIAQIGFREADELDTIVAEGKTKDLARVEMFTADYIDYQAEWARYEGE 122
QY 126 IIGSDRPGNNILFKRALVSTGTITPMNPPFLIARKMAPALITGTITIKSEPTTNA 185
DB 123 IIGSDRPGNNILFKRALVSTGTITPMNPPFLIARKMAPALITGTITIKSEPTTNA 182
QY 186 IAFKIVDEIGLPGVENLVLRGETVGOELAGNPKRVAVVSMTSYSAKEIMATAAKNI 245
DB 183 HIFAEIVDAVGLAGVAVNVNNGAGELGNALSAHPQVDVNSLIGSVEAKQUMELASANI 242
QY 246 TKVLELGGKAPALVMDADILAVKAIYDSRVINSQVCNCAERYVYVKGITIDYFVNU 305
DB 243 TKVLELGGKAPALVMDADILAVKAIYDSRVINSQVCNCAERYVYVKGITIDYFVNU 302
QY 306 GFAGVAVQPCNP--AFENPIAMGFLITMAALIEHVEGVAVPAVPEZAPVAVGKAVGSKGY 363
DB 303 TAAKGVYRGVPAEAGALEMGPILERRVAVKAVKVERAKGKGLVCGGKRAFGRT 362
QY 364 YVPTLLDVKQKSIHMEETFGVPLVPAVFTLEDAISMANSDYGLTSSLYTONLVNA 423
DB 363 FFEPTLLDITNSMTIMKEETFGVPLVPAVFTLEDAISMANSDYGLTSSLYTONLVNA 422
QY 424 MKATIKLKPCTYINKEPEAKMGPFAAGWPKSSIGAGACCKKHJHAGVLCQGVVYVLS 479
DB 423 FYVTPDGFCTYINPEPEAKMGPFAAGWPKSSIGAGACCKKHJHAGVLCQGVVYVLS 478

```

```

RESULT 5
US-10-417-886-5824
: Sequence 5824, Application US/1041/886
: GENERAL INFORMATION:

```

```

: APPLICANT: Keith G. Weinstock et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
: FILE REFERENCE: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
: CURRENT APPLICATION NUMBER: US/10/417,886
: PRIOR FILING DATE: 2003-04-17
: PRIOR APPLICATION NUMBER: US/09/252,691C
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/094,145
: PRIOR FILING DATE: 1998-07-24
: PRIOR APPLICATION NUMBER: US 60/074,787
: PRIOR FILING DATE: 1994-07-18
: NUMBER OF SEQ. ID NOS: 11326
: SEQ. ID NO: 5824
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Enterobacter cloacae
US-10-417-886-5824

```

```

Query Match 61.4% Score 1499; DB 6; Length 319;
Best Local Similarity 61.5% Pctd No. 9 4e-123;
Matches 292; Conservative 17; Mismatches 10; Indels 0; Gaps 0;

QY 161 RKMAPALLTNTIVIMPSEFTNNALAFKIVDEIGLPGVNLVLRGETVGOELAGNP 220
DB 1 RKMAPALLTNTIVIMPSEFTNNALAFKIVDEIGLPGVNLVLRGETVGOELAGNP 60
QY 221 KVAVVMTGSVSAKEIMATAAKNITKVCLEGGKAPALVMDADILAVKAIYDSRVIN 280
DB 61 KVAVVMTGSVSAKEIMATAAKNITKVCLEGGKAPALVMDADILAVKAIYDSRVIN 120
QY 281 SGVCNCAERYVYVKGITIDYFVNRIGEMQAVQFGPARRNDIANGPLINAALEVRQK 340
DB 121 TGGVYVNPFAEVYVQKTYDPFVNPVSEAKVAGCPNPAFTIAPKPLINAAALFVQOK 180
QY 341 VARAVEGAVVAVGKAVGSKGYVYPTLLDVKQKSIHMEETFGVPLVPAVFTLEDA 400
DB 181 VARAVEGAVVAVGKAVGSKGYVYPTLLDVKQKSIHMEETFGVPLVPAVFTLEDA 240
QY 401 ISMANSDYGLTSSLYTONLVNAKAIKIKREYIYNLBNFVAVGKAVGSKGYVY 460
DB 241 ISMANSDYGLTSSLYTONLVNAKAIKIKREYIYNLBNFVAVGKAVGSKGYVY 400
QY 461 DGKGLHGYLOTGVVYVLS 479
DB 301 DGKGLHGYLOTGVVYVLS 319

```

```

RESULT 6
US-10-282-122A-54340
: Sequence 54340, Application US/10/282.122A
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Walli, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282.122A
: PRIOR FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727

```

PRIOR FILING DATE: 1997-08-14

```
9 MYIDGQFEVWFRGDAMITVVNPATEAVISIPYGAETARKAIIOAFPAQPEWEALPAIER 68
```

```

Db      5  LTIDGWNSSSGKTVKSPVIGVIGKFAATRIHIVKRIIUAFAFAMNMLNSVER 64
OY      69  ASMLRKISAGIREPASEISALIVEEGKIQQLAEVAFADYIDYMAEMAKRYEPLIQ 128
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      65  SKIIFYAKELIERNAFLNLIEMENCKPVKAEKEVIGVLIQVUYAEMARKLNGEVE 124
OY      129  SDRPGENILFKRALGVTTGILPNWPFELLARKKAPALITGNITVIRKPSFTNNAIAR 188
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      125  GTSSHHKIFQYKVPYGIYVALITPMWIPDMVARKIAPALLIGNIVLKPSSDFGSAEMI 184
OY      189  AKIVDFIGIPGVFLVIGPGETVQPLACNPKVAMVSMTSVSAEKIMATAKNTKV 248
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      185  VKKFEAGVPKGVNLFTIRGSEIGDYVEHKKNLITMTGSTATGDIRMOKASANAKL 244
OY      249  CLEGGKAPALVMDADLELAVKALVDSRVINSQVNCARVYVQKIQDFYNRLGEA 308
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      245  ITELQKAFWYWKADMDNMLKTLIMAKYNNACOSCIABERLYVHDYIDTFMSRVEL 304
OY      303  MNAVQGNPAEKNDIAMKPLINAAALLEVEKVARAVEGAKVAFGK--AVEG--KGY 363
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      305  SKRLLDGPKNAD--MGPLINKGALQNTSEIVEPAKFSKALIFGSGQPSLQSPYNGY 361
OY      364  YVPPTLLDVRQEMSIHEETGPVLVVAFTLEDAISMANSDYGTSSITYQNLVA 423
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      362  FLPLTIGNADOKSKIPOEPIFAPYIGARKISSVEEMYDILANDSKYGLASTLFTKPNIT 421
OY      424  MKAIGLKFGEIYINRENFAMOGFHAGKRSIGGADGKHGCHGYLOTQVYVL 477
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      422  PEASERIRFEGELVNMFGFEASOGYHTGFRMTGOAGEGSKYCISEYILKNIVY 475

RESULT 9
US-10-282-122A-69547
Sequence 69547, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, P.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-04-20
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2003-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,441
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/267,636
PRIORITY FILING DATE: 2001-02-06
PRIORITY APPLICATION NUMBER: 60/269,408
PRIORITY FILING DATE: 2001-02-16
Remaining Priority Application data removed - See file wrapper or PAM
NUMBER OF SEQ ID NOS: 78614

```

```

: SOFTWARE: PatentIn version 3.1
: SEQ ID NO: 69547
: LENGTH: 480
: TYPE: PRT
: ORGANISM: Pseudomonas syringae
US-10-282-122A-69547

Query Match      30.8%; Score 500.5; Ds 6; Length 480;
Best Local Similarity 39.3%; Pred. No. 5,36-70;
Matches 184; Conservative 101; Mismatches 180; Indels 3; Gaps 2;

OY      16  YTDGQVIVMPLQAMITVNIIVTEAVISPTIFFQVMEFAKATIAAFQAPWFLPAIFPA 69
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      14  YTNQAMLIAMISQSKIKVNNPATINELITVPMGAETIRAEADKALPMAWLTAKFRG 73
OY      70  SWIFKISAGIREPASEISALIVEEGKIQQLAEVAFADYIDYMAEMAKRYEPLIQ 129
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      74  NKLRFWFLMTEINLELUDLLMLLEQKLAIAKKEIITYASPIEMFAEKAPYGGVIG 133
OY      130  DRPGFNILFKPALGVTTGILPNWPFELLARKKAPALITGNITVIRKPSFTNNAIAR 189
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      134  HOPDKRLIVLKOPIGVTAITPMWIPAMITRKGPALAGCTMVLPASQTFPSALATA 193
OY      190  KIVDEIGIPGVENIVLGSGETVGGELAGNPKVAMVSMTSVSAEKIMATAKNTKV 249
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      194  ELAEKAGIPAGVFSVVISAGDISLIGNIVKLSFTSGTEIGKQLMACEKADIKKYS 253
OY      250  EELGSKAPALVMDADLELAVKALVDSRVINSQVNCARVYVQKIQDFYNRLGEAM 309
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      254  TELGGNAPETVFDADLDKAVEGAMISKYRNNGQTCANRIYQDGVYDAFAEKLKVA 313
OY      310  QAVQGNPAERNDIAMKPLINAAALLEVEKVARAVEGAKVAFGKAVGSGYVPPTL 369
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      314  GKELKISGDE DGLIGLEIEKAAKAKHIAVAKCAVLIGSGSLCS--FLIL 370
OY      370  LIDVRQEMSIHEETGPVLVVAFTLEDAISMANSDYGTSSITYQNLVA 429
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      371  LVNKKDAVAAREEFPGLAPLFPRKEAEKIALALANTFPEISAFYFAGMSSVFPVAEA 430
OY      430  LKFGFIYINRENFAMOGFHAGKRSIGGADGKHGCHGYLOTQVYVL 477
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      431  LEYGVVIGINTGLISNELAPFGG1KSSGLGREGSKYGIEDYLEIKYDL 478

RESULT 10
US-10-282-122A-70076
Sequence 70076, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-26
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2003-01-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347

```

```

RESULT 12
Sequence 43455, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Walli, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA-034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/267,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,435
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/240,447

```

```

: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,616
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: SOFTWARE: Patent version 3.1
: SEQ ID NO 43455
: LENGTH: 483
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-43455

```

```

Query Match      36.0%; Score 879.5; DB 6; Length 483;
Best Local Similarity 39.3%; Pred. No. 3,7e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 10 YIDGQFVTWRGDAMIDVVPATEAVISRIIPDGAEDARKAIDAERAPQEMELPATEPA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 YVDGAMVADNDGQITKVNPPATIGETISVPMKGAETRALEADKALPAMKALITAKERA 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 SWLRKISAGIRERASEISALIVEEGKIQQLAEVEVATADYIDYMAEMARPEGETTQS 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 NKLRFWFLMIENQDILARIMTEQSKPLAKAKSHIAYAASTLEMPGEKAKRIYGDPLQS 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 DDPGNNILFKPALVTGTGILPMNPFFLIAPKAPALITGNTIVTPSEFTTNATAPA 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 HOPDKRTIVIKQPIGVTAITPMPNSAMITPKAPALAACTWVLKPAASCTPSALALA 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 KIVDEIGLPRGVFNILVLRGETVGOELAGNPKVAMVMTGVSASGEKIMATAKNITKYC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 ELAEPAAGIPKGVSVVTSAGVSGSLTNSPIVKKLFTSTELGPDLMACADQIKKVS 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 TELGSKAPALYMDADLELAKAIVDSRVINSQVWQWAAHVYVYKQSIYQDFVKKLEAM 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 254 LELGNAPEIYEDDADLAAVEGALISKYPNNGQVCANLYQDQVYDAFVDRKRAAV 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 QAVQGNPAERNDIMQSLINAALEPEYKAVAVAFEGGAVAGGKAVESGYYPPTL 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 AKLNINGNLEAG-VITGPIILAKAKAVAEHIALAVSKAKAVSSKKHALGTFEFTPI 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 LITVPGEMSIHMETFGPVLPVAVFTLEPAISMANSTVYGLTSTYTQNTNVAMKATNS 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 373 LVDVPRKNAIVSKDETFGLAPVFRKLEAEVIAVMSNTERGLASVFAHDLAVFRVAEQ 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 LKHGETYINKENFEAMUGFHAMWKKSSGIGADKHLGSLYQDQVVL 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 433 LEYGMVINTGLISNEVAPFGSIKASJGPRGSKYIEDYLEIKYIDL 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

```

US-10-366-683-27749
: Sequence 27749, Application US/10366683
: GENERAL INFORMATION:
: APPLICANT: Rubenfeld, Marc J.
: APPLICANT: Nolling, Jork
: APPLICANT: Deloughery, Craig
: APPLICANT: Bush, David
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: PAT03-04
: CURRENT APPLICATION NUMBER: US/10/366,683
: PRIOR FILING DATE: 2003-02-13
: PRIOR APPLICATION NUMBER: 03/252,991
: PRIOR FILING DATE: 1999-02-18
: NUMBER OF SEQ ID NOS: 33142

```

```

: SEQ ID NO 27749
: LENGTH: 594
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-10-366-683-27749

```

```

Query Match      36.0%; Score 879.5; DB 6; Length 594;
Best Local Similarity 39.3%; Pred. No. 5,2e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 10 YIDGQFVTWRGDAMIDVVPATEAVISRIIPDGAEDARKAIDAERAPQEMELPATEPA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 YVDGAMVADNDGQITKVNPPATIGETISVPMKGAETRALEADKALPAMKALITAKERA 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 SWLRKISAGIRERASEISALIVEEGKIQQLAEVEVATADYIDYMAEMARPEGETTQS 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 NKLRFWFLMIENQDILARIMTEQSKPLAKAKSHIAYAASTLEMPGEKAKRIYGDPLQS 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 DDPGNNILFKPALVTGTGILPMNPFFLIAPKAPALITGNTIVTPSEFTTNATAPA 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 HOPDKRTIVIKQPIGVTAITPMPNSAMITPKAPALAACTWVLKPAASCTPSALALA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 KIVDEIGLPRGVFNILVLRGETVGOELAGNPKVAMVMTGVSASGEKIMATAKNITKYC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 ELAEPAAGIPKGVSVVTSAGVSGSLTNSPIVKKLFTSTELGPDLMACADQIKKVS 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 TELGSKAPALYMDADLELAKAIVDSRVINSQVWQWAAHVYVYKQSIYQDFVKKLEAM 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 LELGNAPEIYEDDADLAAVEGALISKYPNNGQVCANLYQDQVYDAFVDRKRAAV 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 QAVQGNPAERNDIMQSLINAALEPEYKAVAVAFEGGAVAGGKAVESGYYPPTL 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 AKLNINGNLEAG-VITGPIILAKAKAVAEHIALAVSKAKAVSSKKHALGTFEFTPI 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 470 LITVPGEMSIHMETFGPVLPVAVFTLEPAISMANSTVYGLTSTYTQNTNVAMKATNS 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 LVDVPRKNAIVSKDETFGLAPVFRKLEAEVIAVMSNTERGLASVFAHDLAVFRVAEQ 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 LKHGETYINKENFEAMUGFHAMWKKSSGIGADKHLGSLYQDQVVL 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 544 LEYGMVINTGLISNEVAPFGSIKASJGPRGSKYIEDYLEIKYIDL 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14

```

US-10-419-128-27749
: Sequence 27749, Application US/10419128
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfeld et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196-136
: CURRENT APPLICATION NUMBER: US/10/419,128
: PRIOR FILING DATE: 2003-04-21
: PRIOR APPLICATION NUMBER: US/09/252,391
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 27749
: LENGTH: 594
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-10-419-128-27749

```

```

Query Match      36.0%; Score 879.5; DB 6; Length 594;
Best Local Similarity 39.3%; Pred. No. 5,2e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 10 YIDGQFVTWRGDAMIDVVPATEAVISRIIPDGAEDARKAIDAERAPQEMELPATEPA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 YVDGAMVADNDGQITKVNPPATIGETISVPMKGAETRALEADKALPAMKALITAKERA 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Gen-Pro-Visio 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 24, 2003, 10:10:05, Search time 18.3303 Seconds

(without alignments)
2512 144 Million cell updates/sec

Title: US-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVVVGHMHTIDGQFVIMWS
... AFGKMSLHGYLTGVVYVLS 474

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2432	99.4	479	2 A38165	lactaldehyde dehyd
2	2426	99.2	479	2 E90881	aldehyde dehydroge
3	2426	99.2	479	2 D85737	aldehyde dehydroge
4	1557	63.7	480	2 A81023	aldehyde dehydroge
5	1552	63.5	480	2 F81965	lactaldehyde dehyd
6	1279.5	52.3	393	2 E81394	probable lactaldeh
7	879.5	36.0	483	2 D81613	succinate-semialde
8	873.5	35.7	482	2 F65045	succinate-semialde
9	870.5	35.6	482	2 B91069	succinate-semialde
10	870.5	35.6	482	2 D85213	succinate-semialde
11	858.5	35.1	482	2 AE0844	succinate-semialde
12	845.5	34.6	475	2 D84064	succinate-semialde
13	836	34.2	491	2 C95248	probable succinate
14	835	34.1	468	2 C83774	succinate-semialde
15	822	33.6	482	2 B87648	succinate-semialde
16	821.5	33.6	482	2 A13300	succinate-semialde
17	814.5	33.3	484	2 A95389	succinate-semialde
18	814.5	33.3	484	2 E96825	succinate-semialde
19	811.5	33.2	488	2 A11188	hypothetical prote
20	809	33.1	488	2 A11546	succinate-semialde
21	804.5	32.9	476	2 A74643	glycine betaine al
22	804	32.9	495	2 A75608	aldehyde dehydroge
23	804	32.9	472	2 C74774	succinate-semialde
24	798.5	32.6	486	2 A98308	actK protein (US94
25	798.5	32.6	486	2 A92975	succinate-semialde
26	795.5	32.5	484	2 AE3176	NAD-dependent succ
27	795.5	32.5	490	2 E95316	succinate-semialde
28	795	32.5	494	2 B95411	probable aldehyde
29	791.5	32.4	487	2 C96279	CAN3 succinate-se

RESULT 1

A38165 lactaldehyde dehydrogenase (EC 1.2.1.22) alda - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 01-Mar-2002

C:Accession: A38165; B64893

R:Ridalego, E.; Chen, Y.M.; Lin, E.C.C.; Aguilar, J.

J. Bacteriol. 173, 6118-6123, 1991

A:Title: Molecular cloning and DNA sequencing of the Escherichia coli K-12 ald gene

A:Reference number: A38165; MIMD 92011371; PMID 1917845

A:Accession: A38165

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <HID>

A:Cross-references: GR M64541, NID:J15221, PIR:AA023427, PDB:Q145222

A:Flattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A: Koster, D.J.; Mau, R.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720, MIMD 92126417, PMID 9278503

A:Accession: B64893

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-479 <BLAT>

A:Cross-references: GR AF006239; GR F000926; NID:J1787682; PIR:AA074407.1; PDB:Q17876

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: alda

C:Function: catalyzes oxidation of lactaldehyde to lactate using NAD

A:Note: acts also on other aldehydes

C:Superfamily: aldehyde dehydrogenase (NMD), aldehyde dehydrogenase homology

C:Keywords: NAD, oxidoreductase

F:45-309/Domain, aldehyde dehydrogenase domain

F:178-257/Domain, NAD binding status predicted -NMD

F:251-285/Active site, Gln, Cys #status predicted

Query Match

Best Local Similarity 99.4%, Pred. No. 5, 4e-122

Matches 477: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60

1 MSVVVGHMHTIDGQFVIMWS:PPGQAFPAKAPDAFPAQPM 60


```

Db      121 VVGEELSSNNELGNVSLTGVPAATHTHMAAANKITIKVSLDGGKAPALVCRADITDLAV 180
OY      271 KALVDVSVINSGVGN+AERVYVQKLTIDQFVNRKLGCAQAVQVQGNPAERNDIAMPGLIN 340
Db      181 EAKASHRIQNNQVGN+AFRAVYVHTSVYDEFVQKPEVAKSVSVNTI+KGFEDMGPLVN 239
OY      441 AALAEERQKAVARVEGKAVARVAFQKAVEGKGYYPPTLLVROBMSIMHEETGPPVP 390
Db      240 GAVVGNALMIGVATAKGALVPGQKLTDTSGYFPASVITNKHDEIMKKEITAPILP 299
OY      491 VVAUTLEDAISMANSDVGLTSSITVQNLNVAKAKIKGKGGETYINRENEAMOGPIIA 450
Db      400 IAKUTIDEVIMANKPEYGLTSSITVQNLIDIMRASREIKGGETYINRENEAMOGPIIA 359
OY      451 GMRSGSGTANCKKRGHCHGYLOGVYVLO 478
Db      460 GFRKSSITGGAMKRGHGLEEYLAITHVYVLO 487

```

RESULT 7

```

Db4613 succinate-semialdehyde dehydrogenase PA0265 [imported] Pseudomonas aeruginosa (strain
C)Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: DB4613
C/Author: C.K. Plam, X.Q. Ewin, A.L. Mizoguchi, S.D. Warren, F. Hickey, M.J. Br
adman, S.J. Yuan, Y. Brody, L.L. Coulter, S.N. Forder, K.R. Kas, A. Iatig, K. Iim,
C. Lory, S. Olson, M.V.
Nucleotide: 406, 959,964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: AB2950; MIMD:20437337; PMID:10984043
A/Accession: DB4613
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-483 <STAT>
A/Cross-References: GR:AE004464; GR:AE004091; NID:q9946099, P1DN AAF03664 1, GSPFR-6001
A/Experimental source: strain PA01
C/Genetics:
A/Genes: qabD, PA0265
C/Suprafamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```

```

Query Match      46.08; Score 879.5; Db 2; Length 483;
Best Local Similarity 49.38; Pred. No. 1,2e-53;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;
OY      10 YLKGQVYTWKIDAMIVVNPATFAVTSRTPKQAEIARAKAIIAAEKAPFEMALPAIERA 69
Db      14 YVKAAMVDLNCOTIKVNPATGELIGSVKMKAAETTRATLPAUKALPAMRALTAKEBA 73
OY      70 SWLPKTSATIPPASSTSAIIVHPGKLTQGLAEVAVATVITYMAKARKRYEGTIGS 129
Db      74 NKLRFWEDMTENQDILAKMTLQCKPLAEAKGELAAASPEEMEGEAKKIYDITPG 133
OY      140 DRKGENTLLEKRALGVTTGILWNPPEPLIARKMAFALTGNTIVIKSEFTNNALIFA 189
Db      144 HQRQKRLIVKQPIGTAAITWNPFSAMITRKAGIALAAGCTWKLKASQTFPSALALA 193
OY      190 KIVDELGKGVNLVAKGFTVGOHLACNPKVAVASMTGSASGEKIMATAAKNITKVC 249
Db      194 ELAERATIPKGVSVVTSAGVGELETSPITPKLTFTGSPETRQIMAFADQIKKVS 253
OY      250 LELGKAPVIMVDALIELAVKAVIVSKVINSQVGN+AFRAVYVQKLTIDQFVNRKLGCAQ 409
Db      254 LELGKAPVIVPDADILAEVAGALISKYRNKGQVTCVANKIYVQDGVDAFVQIKAAV 413
OY      410 QAVQGNPAERNDIAMPGLINAAALERVQKAVARVAFQKAVEGKGYYPPTL 469
Db      414 AKININQNGLEAGVTTGPIIDAKAAKAVFPHIADAVSKAKAVSGKRGHALGTFEPTI 472
OY      470 LILVQKPKMSIMHEETGPPVLPVYAFDTLEDAISMANSDVGLTSSITVQNLNVAKAKIKG 429
Db      473 LVLVPPNALVSKDEPGLAPVPERKDEAFVITAMSNDETFGLASVAFADILAPVRAVQ 432

```

```

OY      430 LKRGFTYINPNEFAMQ+FIHAPKRSQIGATPKRHGLHRYIGTVYVYL 477
Db      433 LEYMGVINGLISNEVAPPGIKASGLDEGSKGLIEDYELIKYVYL 480

```

RESULT 8

```

Db5045 succinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) - Escherichia coli (strain
C)Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 03-Jun-2002
C/Accession: F65045
C/Author: F.R. Plunkett III, G. Bloch, C.A. Perna, N.T. Burland, V. Riley, M.
.A. Rose, D.J. Mau, B. Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MIMD:97426117; PMID:9278503
A/Accession: F65045
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-482 <BLAT>
A/Cross-References: GR:AE000361; GR:U00096; NID:q178911; P1DN AAF75708 1; P1D:q17890
A/Experimental source: strain K-12, Substrain MG1655
C/Genetics:
A/Genes: qabD
C/Suprafamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C/Keywords: oxidoreductase

```

```

Query Match      35.78; Score 873.5; Db 2; Length 482;
Best Local Similarity 39.58; Pred. No. 3e-53;
Matches 184; Conservative 91; Mismatches 187; Indels 1; Gaps 1;

```

```

OY      11 IDGQVYTWKIDAMIVVNPATFAVTSRTPKQAEIARAKAIIAAEKAPFEMALPAIERA 70
Db      15 INGMFLDANNGEALIDVTPNANGDKIGSVKMGADETTRATLPAUKALPAMRALTAKEBA 74
OY      71 WLPKTSATIPPASSTSAIIVHPGKLTQGLAEVAVATVITYMAKARKRYEGTIGS 129
Db      75 LKRWPNIMHGHQDLAKMTLQCKPLAEAKGELAAASPEEMEGEAKKIYDITPG 134
OY      131 DRKGENTLLEKRALGVTTGILWNPPEPLIARKMAFALTGNTIVIKSEFTNNALIFA 190
Db      135 QAKRRLIVKQPIGTAAITWNPFSAMITRKAGIALAAGCTWKLKASQTFPSALALA 194
OY      191 LVELGKGVNLVAKGFTVGOHLACNPKVAVASMTGSASGEKIMATAAKNITKVC 250
Db      195 LALRAGVAPGVFVVTQSGAVGQNEITSPILVKLISFTGSEITGQIMEGAKDIKKVSL 254
OY      251 ELGKAPVIMVDALIELAVKAVIVSKVINSQVGN+AFRAVYVQKLTIDQFVNRKLGCAQ 410
Db      255 ELGKNAPVIVPDADILAEVAGALISKYRNKGQVTCVANKIYVQDGVDAFVQIKAAV 414
OY      411 QAVQGNPAERNDIAMPGLINAAALERVQKAVARVAFQKAVEGKGYYPPTL 470
Db      415 KLHIGDGLD+NCVTTGPIIDAKAAKAVFPHIADALFKCAVPGVGGKARPGSNFPQPTL 473
OY      471 LVRQKMSIMHEETGPPVLPVYAFDTLEDAISMANSDVGLTSSITVQNLNVAKAKIKG 430
Db      474 VDPANAKSKSEETPGPLAPLPRFKDEADVLIQANDTFGLAAVYFADLSRVPVQAL 433
OY      431 KPGETYINRENEAMOGPIIAGMPKSGTCAGCKRGLTHRY 471
Db      434 EYGVGNTINGIISNEVAPPGIKASGLDEGSKGLIEDYEL 474

```

RESULT 9

```

Db1069 succinate-semialdehyde dehydrogenase [imported] - Escherichia coli (strain 0157:H7, S
C)Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: B91069
C/Author: T. Makino, K. Ohnishi, M. Kurokawa, K. Ishii, K. Yokoyama, K. Han, C
gasawara, N. Yasunaga, T. Kohara, S. Shiba, T. Hattori, M. Shinagawa, H.

```

DNA Ref. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene

A:Reference: PubMed 119229, M010:21156291, PMID:11208746

A:Accession: F01969

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1,483,144

A:Cross-references: GR:U000007; FIDN:HA00045; FID:0136293; GSPDB:G000154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

A:Gene: ECS3522

C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology

Query Match 35.6% Score 870.5; DB 2; Length 482;

Best Local Similarity 39.3%; Pred. No. 4 9e-53;

Matches 181; Conservative 92; Mismatches 187; Indels 1; Gaps 1;

OY 11 IDGQEVYWRGDAMIDVNPATEAVISRIIDGQAFPAKPAIDAEAPQEFALPAIRAS 70
 DB 15 INCEMLDANNGEVIDVTPNDKDKISVPMKCADETRALDAANRALPVMALIAKERAN 74
 OY 71 WLKRTISAGIREBASEISALIVEEGCKIQQLAEVAVTADYDYMAEMARPYEGEITQSD 130
 DB 75 ILRNMENLMMEHQDLARLMTLEQSKPLAEAKGRTSYAASFTEWFAEGRKPTGDTTIGH 134
 OY 131 RGENILLFKRALGVTTGILPWNPPFLIARKMAPALLTGNTIYIKSEFTNNALAFK 190
 DB 135 QADKRLIVKOPTVTAITTPNPPAMITRKAGALASQVIMWIKRPSQTPSALALAE 194
 OY 191 IYDEIGLPGVENVLVSGETVGOELAGNPKVAVSMGVSAGEKIMATAKNITKVL 250
 DB 195 LAIPAGTIPAGVENVVTSAGAVNGEITSNPLVRLSTFETISLMLPHQCAKDICKVSL 254
 OY 251 ELGKAPALVMDALLELAVKAVISVNSQVNCNCKEYVYKGIYDQVNLGSAQ 310
 DB 255 ELGSAAPFTVFDALDCAVEBQALASKRPNALQTCVANKLYVQGVYDRPAEKLDQAVS 314
 OY 311 AVQGNPAERNDIAMSPILMAALEEYBQKVAFAVEBQKVAFAVEBQKVAFAVEBQKVA 370
 DB 315 KLHIGGLDKG-VTIGPLIDEKAVAKVEBHADALEKARAVCGKAKERGNFOTIL 373
 OY 371 LDVROBMSIMHEETGPVLPVVAFTLEDAISMANDSYGLTSSITONLVANKAIKGL 430
 DB 374 VVPANAKVSKETGGLAPLRFEPDEADYIAQNDPEFGIAAFYAFILSRVPRVEAL 433
 OY 431 KPGFTYINPENPEAMQFPAKWPKSGIGADGKHLHGYLQ 471
 DB 434 EYGIQVINGTILISNEVAPRGKIKASGLRGSKTIGEDYLE 474

RESULT 10

D85913

succinate-semialdehyde dehydrogenase [imported] - *Escherichia coli* (strain O157:H7, subs

C:Species: *Escherichia coli*

C:Date: 10-Feb-2001 #sequence-revision 16 Feb 2001 #text-change 27-Nov-2001

C:Accession: D85913

R:Perma. N.T. Plunkett III, G. Nuland, V. Muir, R. Glesner, J.P. Rose, D.J. Mayhew

iller, L.J. Grobeck, E.J. Davis, N.W. Lim, A.J. Dimalanta, E.J. Polomousis, K.J. Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; M010:21074935; PMID:11208551

A:Accession: D85913

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1,483,144

A:Cross-references: GR:AE005174; NIT:012517054; FIDN:AA157768; GSPDB:G000145; U000007

A:Experimental source: strain O157:H7, substrain EDL933

C:Gene: 93bD

C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology

Query Match 35.6% Score 870.5; DB 2; Length 482;

Best Local Similarity 39.3%; Pred. No. 4 9e-53;

Matches 181; Conservative 92; Mismatches 187; Indels 1; Gaps 1;

OY 11 IDGQEVYWRGDAMIDVNPATEAVISRIIDGQAFPAKPAIDAEAPQEFALPAIRAS 70
 DB 15 INCEMLDANNGEVIDVTPNDKDKISVPMKCADETRALDAANRALPVMALIAKERAN 74
 OY 71 WLKRTISAGIREBASEISALIVEEGCKIQQLAEVAVTADYDYMAEMARPYEGEITQSD 130
 DB 75 ILRNMENLMMEHQDLARLMTLEQSKPLAEAKGRTSYAASFTEWFAEGRKPTGDTTIGH 134
 OY 131 RGENILLFKRALGVTTGILPWNPPFLIARKMAPALLTGNTIYIKSEFTNNALAFK 190
 DB 135 QADKRLIVKOPTVTAITTPNPPAMITRKAGALASQVIMWIKRPSQTPSALALAE 194
 OY 191 IYDEIGLPGVENVLVSGETVGOELAGNPKVAVSMGVSAGEKIMATAKNITKVL 250
 DB 195 LAIPAGTIPAGVENVVTSAGAVNGEITSNPLVRLSTFETISLMLPHQCAKDICKVSL 254
 OY 251 ELGKAPALVMDALLELAVKAVISVNSQVNCNCKEYVYKGIYDQVNLGSAQ 310
 DB 255 ELGSAAPFTVFDALDCAVEBQALASKRPNALQTCVANKLYVQGVYDRPAEKLDQAVS 314
 OY 311 AVQGNPAERNDIAMSPILMAALEEYBQKVAFAVEBQKVAFAVEBQKVAFAVEBQKVA 370
 DB 315 KLHIGGLDKG-VTIGPLIDEKAVAKVEBHADALEKARAVCGKAKERGNFOTIL 373
 OY 371 LDVROBMSIMHEETGPVLPVVAFTLEDAISMANDSYGLTSSITONLVANKAIKGL 430
 DB 374 VVPANAKVSKETGGLAPLRFEPDEADYIAQNDPEFGIAAFYAFILSRVPRVEAL 433
 OY 431 KPGFTYINPENPEAMQFPAKWPKSGIGADGKHLHGYLQ 471
 DB 434 EYGIQVINGTILISNEVAPRGKIKASGLRGSKTIGEDYLE 474

RESULT 11

AE0839

succinate-semialdehyde dehydrogenase (NAD(P)) (ec 1.2.1.16) - *Salmonella enterica sub*

C:Species: *Salmonella enterica* subsp. *enterica* serovar *Typh*

A:Note: this species has also been called *Salmonella typhi*

C:Date: 09-Nov-2001 #sequence-revision 09-Nov-2001 #text-change 03-Jun-2002

C:Accession: AE0839

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.P.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, P.M.; Dowd, L.; White, N.; Fair

clough, S.; Mouton, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se

A:Reference number: AB0502; PMID:11677608

A:Accession: AE0839

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1,482,749

A:Cross-references: GR:AL11342; FIDN:AA05000; FID:0136293; GSPDB:G000176

A:Gene: STY2911

C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology

C:Keywords: oxidoreductase

Query Match 35.1% Score 858.5; DB 2; Length 482;

Best Local Similarity 39.2%; Pred. No. 4 4e-52;

Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

OY 10 YHGGQEVYWRGDAMIDVNPATEAVISRIIDGQAFPAKPAIDAEAPQEFALPAIRAS 69
 DB 14 YHGGQEVYWRGDAMIDVNPATEAVISRIIDGQAFPAKPAIDAEAPQEFALPAIRAS 73
 OY 70 WLKRTISAGIREBASEISALIVEEGCKIQQLAEVAVTADYDYMAEMARPYEGEITQSD 129
 DB 74 WLKRTISAGIREBASEISALIVEEGCKIQQLAEVAVTADYDYMAEMARPYEGEITQSD 133
 OY 130 RGENILLFKRALGVTTGILPWNPPFLIARKMAPALLTGNTIYIKSEFTNNALAFK 189

Db 134 HQTDRRLVLTGKPLVIAITLTPWNPISAMITKRNALALAGCTMTVLRKASQTPSALIA 193
 Oy 190 KIVETLLRGGVNLVLPSPETVGGELACNPKVAMSTGVSACHIKMATAKNITVGC 249
 Db 194 ELANRACILAGVNNVVGASADIGGELTNPVLRKLSPTGSTEIGKQIMEGCAKDKRVS 253
 Oy 250 LELAKKAVLVMDDADELAVKAIYDSRVINSQVNCAPRYVOKGIDYQVNRIGEM 309
 Db 254 LELAGNAPITVTDADILKAVRGALASKFRNAGCTVCANRKYVDQVYDRFAKRLNAV 313
 Oy 410 QAVQNPAPERNIIMQPIINAAALERVQKAVARVEGAKVAFSGKAVEGKYPPTL 369
 Db 414 NKLAVRGLQGLVATGPELDEKAVAKVGEHLADALEKGRVITGGEAKRLGNGEPTT 372
 Oy 470 LLIVRQFMSIMHEETGGVLPVVAFTLLEDAISMANDSYGCTSSITYONLNVAKAKIG 429
 Db 473 LAIVPDKAKAKKEETGPAFLERESDEADVLRQAMTTEFGIAHYVADHISPVVQCPA 432
 Oy 440 LKREETYINRENEAMQGHMGPSSGIGAPVCKHGLGTYLQ 471
 Db 443 LEGLVGLINTGLISNEVAPPGCIKASGLGKESKYEITYLE 474

RESULT 12
 Db4064
 sucinate-semialdehyde dehydrogenase gabd [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 *sequence_revision 01-Dec-2000 *text_change 15-Jun-2001
 C:Accession: DB4064
 R:Tokami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
 Nucleic Acids, Res. 29, 4417-4431, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: AB4650; MIMD:20512582; PMID:11058132
 A:Accession: DB4064
 A:Status: preliminary
 A:Molecule type: DNA
 A:Keywords: 1-475, str.,
 A:Cross references: GB:AF00151b; GB:BA000004; NID:q10175/94; P10N:BAH0705.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genes:
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 44.68; Score 845.5; Db 2; Length 475;
 Best local similarity 39.18; Prod. No. 2.7e-51;
 Matches 184; Conservative 84; Mismatches 195; Indels 7; Gaps 4;

Oy 9 MTIVQFVIMRCIA-WIVVNPATFAVLSRLPDGAPARPKATIAEPAQPEWPAI 66
 Db 6 LYLIG---TWGILLDTFVKNPATGEVYGVMPNGKAEAAALEAQAQAFIKRKYTAA 62
 Oy 67 ERASWIKRISAGIRKASISALIVRECKTQGLAVVVAFTADYIDYMAFWAPPYGEI 126
 Db 63 ERANPIMTMHDLLOCKKEELAEIMTEMCKPLAFKGLLEYASPIPMFAEGSKVYERT 122
 Oy 127 IGDURGENILFKRALGVTTGLIPNPFELIARKMAPALLTGMTIYIKSEPTTNAI 186
 Db 123 VASASNRKTHVQKQVGVAAITPMNPFAMIVKMAALAAAGCTFEGKDELPLIPLTA 182
 Oy 187 AFAKIVDEIGLRGVNLVLPSPETVGGELACNPKVAMSTGVSACHIKMATAKNIT 246
 Db 183 KIVELDEEAFKRGVNLVLPSPETVGGELACNPKVAMSTGVSACHIKMATAKNIT 242
 Oy 247 KIVELDEEAFKRGVNLVLPSPETVGGELACNPKVAMSTGVSACHIKMATAKNIT 246
 Db 243 NISLELQVQPMITDADILKAVRGALASKFRNAGCTVCANRKYVDQVYDRFAKRLNAV 313
 Oy 407 EAMQVQNPAPERNIIMQPIINAAALERVQKAVARVEGAKVAFSGKAVEGKYPPTL 369
 Db 403 GKAVQVQKAVRGLQGLVATGPELDEKAVAKVGEHLADALEKGRVITGGEAKRLGNGEPTT 372
 Oy 466 PTLTLLVQFMSIMHEETGGVLPVVAFTLLEDAISMANDSYGCTSSITYONLNVAK 425

Db 362 DPTLLVDHDEMLVMOELTFGVAPVQTFATGDEVIDEAFANCTRGCLAVFTENTARGLP 421
 Oy 426 ALKQKRGVYINRENEAMQGHMGPSSGIGAPVCKHGLGTYLQ 471
 Db 429 LSPALDPGIVGMNCAFPSIAQAFPGSKESGIAFRVQGGELFALETFKV 471

RESULT 13
 C95948
 Probable sucinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) [imported] - Sin
 C:Species: Sinorhizobium meliloti
 C:Date: 24 Aug 2001 *sequence_revision 24 Aug 2001 *text_change 03 Jun 2002
 C:Accession: C95948
 R:Finan, T.M.; Weidner, S.; Wong, K.; Bihmester, J.; Chain, P.; Vorholter, F.J.; Her
 Proc. Natl. Acad. Sci. U.S.A. 98, 9899-9894, 2001
 A:Title: The complete sequence of the 1.68-kb psymb megaplasmid from the N2-fixing e
 A:Reference number: A95842; MIMD:21396508; PMID:11481431
 A:Accession: C95948
 A:Status: preliminary
 A:Molecule type: DNA
 A:Keywords: 1-491, KURP,
 A:Cross references: GB:AL521985; P10N:CAQ4925.1; P10N:Q15140747; GSPDB:GN0167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hub
 pel, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Jellau
 hehl, P.; Vandenhol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MIMD:21368234; PMID:11474104
 A:Contents: annotation
 C:Genes:
 A:Gene: gabd2; SMD21185
 A:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 34.28; Score 816; Db 2; Length 491;
 Best local similarity 39.18; Prod. No. 1.3e-50;
 Matches 176; Conservative 85; Mismatches 185; Indels 2; Gaps 2;

Oy 25 DVVNPATFAVLSRLPDGAPARPKATIAEPAQPEWPAI 66
 Db 36 DVNPSTGILLATLPMGIDDAKTAIDAAALQPLMAPKPAKDSIILRWHDILVHAD 95
 Oy 85 EISALIVEERGKTOGLAEVEVAFTADYIDYMAFWAPPYGEI 126
 Db 96 DLVALITAEKGRPVGAKGEVLHMAASYEMVDEEAKRYVGTFFAPANDRMILYIKQVVG 155
 Oy 145 VTGGLPMPPEFLIARKMAPALLTGMTIYIKSEPTTNAI 186
 Db 156 VVGTITPMNPFASVARKISPPALAGCTVLEKPAEGOTPLVAGAMPVLAEEKAGFEGVLT 215
 Oy 205 V-LGRGEGVGOELAGCPKAVNSMTGVSAGKIMATAKNITVGCLEGGAKAATYMD 263
 Db 216 LVASGAPVIGRELQGPVKKLSPTGSTEIGKQIMEGCAKDKRVS 253
 Oy 264 ADLEAVKAIYDSRVINSQVNCAPRYVOKGIDYQVNRIGEM 309
 Db 276 ADIDEAVDCAVQAKFPNAGCTVCANRKYVDQVYDRFAKRLNAV 313
 Oy 324 AMQVQNPAPERNIIMQPIINAAALERVQKAVARVEGAKVAFSGKAVEGKYPPTL 369
 Db 335 ALGPMIDAHADIKTANVADAAVAKQAVSGSGRIIGTGTFPEPTLVGLSHMRKIAQPE 394
 Oy 384 TEGVLPVVAFTLLEDAISMANDSYGCTSSITYONLNVAKAKIG 429
 Db 395 TROPVAPITRFTARQVAFANDITYGLAAHYVADHISPVVQCPA 432
 Oy 444 AMQVQNPAPERNIIMQPIINAAALERVQKAVARVEGAKVAFSGKAVEGKYPPTL 369
 Db 455 SFAPDPGIGKSGIGSPGTSRPHGLFDTLF 482

RESULT 14

C83774

succinate-semialdehyde dehydrogenase BH0995 [imported] - *Bacillus halodurans* (strain C-3)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: C83774

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, P.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: C83774

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1468 <STO>

A:Cross-References: GB:AF001510; GB:BA000004; MID:31017440; PIRN:BA004714.1; CSPIR:GN00

A:Experimental source: strain C-125

C:Genetics:

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match

Best Local Similarity 34.1%; Score 835; DB 2; Length 468;

Matches 170; Conservative 105; Mismatches 186; Indels 6; Gaps 3;

9 MYIDGFTWGDAMIDVNPATPAVLSRIPDGAEDARKAIDAERAOPFEALPAIER 68

3 LYINGEMW--RSGKTLDVNPATGEVIDIVPFAGKKAEALVSAAYEAAPQWSQVASIK 60

69 ASWLKRTISAGIRERASEISALIVEEGKIQULAEVEVAFADYIDVMAEWARRYEGELIU 128

61 SPYLMRWQQLIDPDODELGEIMTEKEGKPLREAFGEVQVANSFLQWAEAKRIYGDITP 120

129 SDREGENILLKRALCVTTGILPWNPPFLIARKMAPALLTQNTIYIKSEPTTNNALAF 188

121 ASATIKRRLVOKOPGVAAITPWNPPFAMITRKVAALAGCAIYVPAEQPTPLAIKL 180

189 AKIVDEIGLPRGVNVLVIGRGETVGOELAGNPKVAMVSMGVSAGEKIMATAKNTYK 248

181 AOLAEAGIPAGVNLVITGNMODIGEMLEDSRKRTITFTGSTEYGLLMRGAAYVKKI 240

249 CLEIGKAPALVMDADLELAKAIVDSRVAINSGQVNCARVYVKGIVDQFVNRLGEA 308

241 STEIGGHAPETIIMDANILEAVDQVLAKEFNAGQTVVAFNRYVAFIAPETKFAAK 300

309 MGAVQFGNPAERKNLIAMGHLINAAALRKVQKVAARAEAGARAFGKAAGEGYYRPT 368

301 VNEIKVNGGLEFG-VTTGPIIDKAAEKFVFAHIDALIKKGGQVTVGSP--WTNHEFEPT 356

369 LLLDVROEMSIHHEETGCVLPVPAFPLEDAISMANDSYGLTSSITYTONLVAMKALK 428

357 IITGATDEMILCMNETGEPPLAPVATPDTBEHVIERANHTPYGLAAYFTENIGRAIRLS 416

429 GLKSEETITNPNFEAMQGFHAKWPKSGTGGADSKHGLHYLOTQV 475

417 KLEVGIVGVNDMPSSVAQAPFGWKESGLGREGKYGILEEYLVKKV 463

RESULT 15

B87638

succinate-semialdehyde dehydrogenase [imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: B87638

R:Nierman, W.C.; Fajkiewicz, T.V.; Paulsen, I.T.; Nollson, K.E.; Eisen, J.; Heidelberg, J.

B.; Lamb, M.T.; Derby, R.T.; Dodson, J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87638

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1482 <STO>

A:Cross-References: GB:AE005773; MID:313424806; PIRN:AAK25102.1; CSPIR:GN00148

C:Genetics:

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match

Best Local Similarity 33.6%; Score 822; DB 2; Length 482;

Matches 185; Conservative 82; Mismatches 199; Indels 4; Gaps 3.

5 VHHKLVLDGQVNTWKGAMILVNPATPAVLSRIPDGAEDARKAIDAERAOPFEALPAIER 64

7 VETALLIDQWV--RSGKTLDPATGEVIDIVPFAGKKAEALVSAAYEAAPQWSQVASIK 64

65 ALEASWLRKISASIREKASHISALIVEEGKIQULAEVEVAFADYIDVMAEWARRYEGELIU 124

65 ALEPAPILIPWSTIILAHADHAPLMTIPQKKPLAFAKKEPVVYGASFTIWPFAEAPPAV 124

125 FIIQSDPRENILLFKALVTTGILPWNPPFLIARKMAPALLTQNTIYIKSEPTTNNALAF 184

125 HIIPTWPKKRLASIKQVGVGCAALAPWNPFLAMIIKKVGPALAAKTVVKKPAALPLS 184

185 AIAFAKIVDEIGLPRGVNVLVIGRGETVGOELAGNPKVAMVSMGVSAGEKIMATAK 243

185 ALATAPLATENKCVAGVNLVITTPRSSEVQKVLTPDPSVPRKLSFTGSTPTGKVIYQV 244

244 NITVNDIGLQKALAVIMHLELAVKAVIVSIVNSQVNCARVYVKGIVDQFVNRLGEA 303

245 TMKRLSTELGNAEPIVEEDADLEAVDGAIAKRYRNAGQTCANRLIVQSGIHDAFAA 304

304 KIKAMAVFGNPAERNDIAMPELINAAALFVRLKVAARAEAGARAFGKAAGEGYYRPT 363

305 RLAEVVALKVG--PQTEGVQDGLPLNEKALIKVGLVSAVYAGAKVILGSDVHGLGCH 363

364 YTPPTLLDVROEMSIHHEETGCVLPVPAFPLEDAISMANDSYGLTSSITYTONLVAMK 423

364 FYQPTVLVAGATPEKRIFOEELIFGVADIVKFEAEVVELANATPGSLAAYEYSRDVGC 423

424 MKATKSKIKFSTYTNPNFAMQGFHAKWPKSGTGGADSKHGLHYLOTQV 473

424 WKVAEQIEAGVNGIEGLISTEVAIPGGVKGESGLGREGKYGILEEYLVKKV 473

Search completed: June 24, 2004, 10:18:31
Job time: 19.3303 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw mode

Run on: June 24, 2003, 10:02:55, Search time: 6.4373 seconds
(without alignments)
2059,300 Million cell updates/sec

Title: us-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVPVQHPHMLDGFVTRG... ADKGHLGVLDTQVLYLOS 479

Scoring table: BIOSUM62
Gap: 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2427	99.2	479	ALDA_ECOLI	P25553 Escherichia
2	874.5	36.7	482	CABD_ECOLI	P25556 Escherichia
3	810	33.1	491	GABD_HISN	P55653 Rhizobium s
4	804.5	32.9	490	PHAR_RACSN	P71016 Bacillus su
5	783.5	32.0	474	YDCW_ECOLI	P77674 Escherichia
6	777	31.8	519	YDCW_ECOLI	P46367 Saccharomyc
7	769	31.4	500	DHAL_YEAST	O74187 agachatus bl
8	767	31.4	488	DHAL_AGABI	P51650 ratius novy
9	759.5	31.1	535	SSDH_RAT	P51649 homo sapien
10	750	30.7	519	DHAL_YEAST	P40047 saccharomyc
11	728	29.8	498	DHAL_LETTA	Q25417 Ichthmania
12	726	29.7	497	DHAL_EMENT	R08157 emeticellia
13	719.5	29.4	512	DHAL_HUMAN	P47895 homo sapien
14	719	29.4	499	DHAL_YEAST	P54115 saccharomyc
15	718.5	29.4	497	CHAB_STFPL	P17202 Stenobacill
16	713	29.1	902	FTCH_RAT	P28037 ratius novy
17	710	29.0	500	DHAL_BFTVU	P28237 ratius novy
18	708.5	29.0	487	DHAL_RHIME	P54232 rhizobium m
19	708.5	29.0	497	UCAS_YEAST	P38067 saccharomyc
20	708	28.9	489	DHAL_ECOLI	P17445 Escherichia
21	708	28.9	497	DHAL_ASPIG	P41761 aspergillus
22	702.5	28.7	493	DHAL_HUMAN	P41899 homo sapien
23	699.5	28.6	503	DHAL_GADCA	P56533 gadus calla
24	699	28.6	902	FTDH_HUMAN	P55891 homo sapien
25	698	28.6	502	DHAL_ATPHO	P42757 atirplex ba
26	696	28.4	500	DHAL_STHPE	P59808 schistosom
27	694	28.4	507	DHAL_RAT	P51647 ratius novy
28	690.5	28.2	497	DHAL_ENTBU	Q27640 booby:ratu
29	687.5	28.1	500	DHAL_MESAU	P81178 mesocricetu
30	687.5	28.1	519	DHAL_MOUSE	P47738 mus musculu
31	686.5	28.0	519	DHAL_RAT	P11884 ratius novy
32	685.5	28.0	484	DHAL_BACSU	P42236 bacillus su
33	684.5	28.0	500	DHAL_HORSE	P15417 equus cabal

34	684	28.0	500	DHAL_MOUSE	P24549 mus musculu
35	683.5	27.9	500	DHAL_SHEEP	P51977 ovis aries
36	681	27.8	496	DHAL_GIAHE	P40108 ciadospoiti
37	680.5	27.8	500	DHAL_HUMAN	P00352 homo sapien
38	678.5	27.7	500	DHAL_BOVIN	P48644 bos taurus
39	678	27.7	500	DHAL_PAT	P13601 ratius novy
40	677.5	27.7	517	DHAL_HUMAN	P05091 homo sapien
41	676.5	27.6	500	DHAL_H-RSP	P12762 equus cabal
42	675	27.6	501	DHAL_APAFH	Q48795 arachidopis
43	675	27.6	501	DHAL_MARCP	Q24490 macroscelid
44	672	27.5	501	DHAL_ELEED	Q48399 elephanthin
45	671.5	27.5	499	FEAR_ECOLI	P06568 Escherichia

ALIGNMENTS

RESULT 1

ID	ALDA_ECOLI	STANDARD:	PRT:	478 AA
AC	P25553:			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Aldehyde dehydrogenase A (EC 1.2.1.22) (Lactaldehyde 3-hydrogenase).			
GN	ALDA OF ALD OF B415.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.			
RC	STRAIN-K12:			
RX	MEDLINE=92011371; PubMed=1917845;			
RA	Hidalgo E., Chen Y.-M., Lin E.C.C., Aguilar J.;			
RT	Molecular cloning and DNA sequencing of the Escherichia coli K-12			
RT	ald gene encoding aldehyde dehydrogenase.*;			
PL	J. Bacteriol 173:6118-6123(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
FX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,			
RA	Pilley M., Collado-VIDES J., Glasner J.P., Rode C.K., Mayhew G.F.,			
RA	Gregor T., Davis N.W., Kirkpatrick H.A., Gordon M.A., Rose P.J.,			
RA	Mau B., Shao Y.;			
RT	The complete genome sequence of Escherichia coli K 12.*;			
RL	Science 277:1453-1474(1997).			
RP	[3]			
RC	SEQUENCE FROM N.A.			
RX	MEDLINE=97251357; PubMed=9037033;			
RA	Alta H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,			
RA	Itch T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,			
RA	Kikugawa M., Makino F., Miki T., Minohara K., Mori H., Maki T.,			
RA	McIntyre K., Nakada S., Nakamura Y., Nishimura H., Nishio Y.,			
RA	Oshima T., Saito J., Sampel G., Seki Y., Sivasubraman S.,			
RA	Tajima H., Takeda J., Takeuchi K., Takeuchi Y., Wada C.,			
RA	Yamamoto Y., Horinouchi T.;			
RT	"A 570-kb DNA sequence of the Escherichia coli K 12 genome			
RT	corresponding to the 28.0-49.1 min region on the linkage map.*;			
RL	PNA Res. 3:363-377(1996).			
RN	[4]			
RP	SEQUENCE OF 1-11.			
RC	STRAIN K12 / EMCC.			
FX	MEDLINE=97443975; PubMed=9298646;			
RA	Link A.T., Robison K., Church G.M.;			
RT	"Comparing the predicted and observed properties of proteins encoded			
RT	in the genome of Escherichia coli K 12.*;			
RL	Electrophoresis 18:1359-1363(1997).			
CC	FUNCTION: ACTS ON LACTALDEHYDROGENASE AS WELL AS OTHER ALDEHYDES.			
CC	CATALYTIC ACTIVITY (S)-lactaldehyde + NAD(+) -> H(2)O + (S)-			
CC	lactate + NADH.			

```

CC -1 SUBUNIT: HOMOTETRAMER.
CC -1 INDUCTION: BY GROWTH ON FUCOSE, RHAMNOSE, APARINOSE AND AMINO
CC ACIDS SUCH AS GLUTAMATE.
CC -1 SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64541; AAC23427.1; -
DR EMBL: AF000249; AAC74497.1; -
DR EMBL: D90780; BAA15032.1; -
DR EMBL: D90781; BAA15037.1; -
DR PIR: A48165; A48165; -
DR HSSP: P51977; 18XS; -
DR SWISS_2DPAGE: P25553; COL1; -
DR EC00900: EC10045; ALD4; -
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd. 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR CAZymeDatabase: NADP; Complete proteome.
DR INIT_MET 0 0
DR NP_BIND 206 212 NAD (ADP PART) (BY SIMILARITY).
DR ACT_SITE 250 250 BY SIMILARITY.
DR ACT_SITE 284 284 BY SIMILARITY.
DR SEQUENCE 478 AA; 52141 MW; 0950F44B29E87F4A CRC64;
Query Match 99.4%; Score 2427; DB 1; Length 478;
Host Label: Similarity 99.6%; Proc. No. 1; 96 155;
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC 2 SVFVQHHMYIDGQFVTRKQAMIDVNVNPAFTEAVISIKIPGQEDARKAIDAERKOPME 61
CC 1 SVFVQHHMYIDGQFVTRKQAMIDVNVNPAFTEAVISIKIPGQEDARKAIDAERKOPME 60
CC 62 ALPAIFKASWIKRISAGIRKASRSEISALIVEEGKIQDLAEVEVAFTADYIDYMAEMARR 121
CC 61 ALPAIFKASWIKRISAGIRKASRSEISALIVEEGKIQDLAEVEVAFTADYIDYMAEMARR 120
CC 122 YEEGIIQSDRPGENILFEKRALGVTTGILPMNPFPELLARKKAPALLTGNTIVIKPSEFT 181
CC 121 YEEGIIQSDRPGENILFEKRALGVTTGILPMNPFPELLARKKAPALLTGNTIVIKPSEFT 180
CC 182 TNNAIAFAKIVDEIGLPGVFNLYIGRGTGQGLAGNPKVAMVMTGTSVSGEKIMATA 241
CC 181 PNNALAFAKIVDEIGLPGVFNLYIGRGTGQGLAGNPKVAMVMTGTSVSGEKIMATA 240
CC 242 AKKIVTVELEGGKAAIVAKIAGLGLAVKAVIVSKVINSGVCMZAKVYVVKCTYQGF 301
CC 241 AKKIVTVELEGGKAAIVAKIAGLGLAVKAVIVSKVINSGVCMZAKVYVVKCTYQGF 300
CC 302 VNRILGEAMQAVQGFNFAERNILAMGPIINAAALEVEVEKVAVAEVEGARVAFGKAVGK 361
CC 301 VNRILGEAMQAVQGFNFAERNILAMGPIINAAALEVEVEKVAVAEVEGARVAFGKAVGK 360
CC 362 GYVYPTLLLDVQKENSIMHEETFCVLTVAVDTLEAISMANDSDYGLTSSITYQNLN 421
CC 361 GYVYPTLLLDVQKENSIMHEETFCVLTVAVDTLEAISMANDSDYGLTSSITYQNLN 420
CC 422 VAMKAIEKIKPGCEVYNRENTFAMGFTIAGGKSKISGACGKHGICVLTQVQVYIQS 479
CC 421 VAMKAIEKIKPGCEVYNRENTFAMGFTIAGGKSKISGACGKHGICVLTQVQVYIQS 478

```

```

DR 01-MAY-1992 (Rel. 22, Created)
DR 01-MAY-1992 (Rel. 22, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DR Succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16) (SSDH).
DR GAAD OR h2661.
DR Escherichia coli.
DR Escherichia coli.
DR Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
DR Escherichia.
DR NCBI_Taxid=562;
DR [1]
DR SEQUENCE FROM N.A.
DR STRAIN-K12 / JM103;
DR MEDLINE=94127927; PubMed=8297211;
DR Niegemann E., Schulz A., Bartsch K.;
DR "Molecular organization of the Escherichia coli gcb cluster:
DR nucleotide sequence of the structural genes gcbH and gcbP and
DR expression of the GABA permease gene."
DR Arch Microbiol 160:454-460(1993).
DR [2]
DR SEQUENCE FROM N.A.
DR STRAIN-K12 / MG1655;
DR MEDLINE=97426617; PubMed=9278503;
DR Blatter F.P., Plunkett G., III, Bloch C.A., Petra N.T., Harland V.,
DR Riley M., Collado-Vides J., Glasner J.D., Fode C.K., Mayhew G.F.,
DR Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
DR Mau B., Shao Y.;
DR "The complete genome sequence of Escherichia coli K-12."
DR Science 277:1233-1238(1997).
DR [3]
DR SEQUENCE FROM N.A.
DR STRAIN-K12;
DR MEDLINE=97349980; PubMed=9205837;
DR Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
DR Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
DR Mizubuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
DR Oshima H., Oiyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,
DR Tagami H., Takahashi H., Takeeda J., Takemoto K., Uehara K., Wada C.,
DR Yanagita S., Horinouchi T.;
DR "Construction of a contiguous 874-kb sequence of the Escherichia coli
DR K-12 genome corresponding to 50,000,000 bp on the linkage map and
DR analysis of its sequence features."
DR DNA Res. 4:91-113(1997).
DR -1- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O =
DR succinate + NAD(P)H.
DR -1- PATHWAY: 4-aminobutyrate (GABA) degradation.
DR -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
DR in positions 51 and 190.
DR -----
DR This SWISS-PROT entry is copyright. It is produced through a collaboration
DR between the Swiss Institute of Bioinformatics and the EMBL outstation -
DR the European Bioinformatics Institute. There are no restrictions on its
DR use by non-profit institutions as long as its content is in no way
DR modified and this statement is not removed. Usage by and for commercial
DR entities requires a license agreement (See http://www.isb-sib.ch/announce/
DR or send an email to license@isb-sib.ch).
DR -----
DR EMBL: M68334; AAC36831.1; -
DR EMBL: AF000051; AAC75708.1; -
DR EMBL: D90890; BAA16522.1; ALT_FRAME.
DR EMBL: D90890; BAA16523.1; ALT_FRAME.
DR EMBL: D90890; BAA16524.1; ALT_FRAME.
DR HSSP: P51977; 18XS; -
DR EC00900: EC11329; gcbH.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd. 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR CAZymeDatabase: NADP; Complete proteome.
DR NP_BIND 233 238 NAD(P) (ADP PART) (BY SIMILARITY).
DR ACT_SITE 255 255 BY SIMILARITY.
DR ACT_SITE 289 289 BY SIMILARITY.
DR SEQUENCE 482 AA; 51720 MW; 091538F8741B0D7C CRC64;

```

Query Match 35.78; Score 873.5; DP 1; Length 482;
 Best Local Similarity 79.54; Pred. No. 2; Gaps 1;
 Matches 182; Conservative 91; Mismatches 187; Indels 1; Gaps 1.

11 IDGGFTWRDAMIDVNPATFAVISRIPDGGAEEDARKKALDAERKQPEWMLPAIERAS 70
 15 INGEWLLANNGEALIDYINFRANGKLSVKKMNAFTPAITDAANRAIPAMRAITAKEPAT 74
 71 WLPKISATFEFASSETALIVEETSTTQIAFEVAVAFADYIIVYMAFMAFVEDELLQSD 130
 75 ILRNFMILMMEHODLAKMTLEQCKPLAEKGEISYASAFTEWFAEDGKRIYGDPIGH 134
 QY 131 KRGNNILFKKALGVITIGLPMFPEFLAKKMAKALIGNIVIKRSEFTINNAIAAK 190
 135 QADKRLIVIKOPIGVITAIIPMNPAMITIRKAGPALAGCTIVLKPASQTFPSALALAE 194
 QY 191 IVDEIGLPRGVNLVLRGETVGOELACGNPKVAMVSMGTSVSAGFKIMATAKNITKVTI 250
 195 LAIFAGVAGVFNVTGSAGVGNELSNPLVKLSFTGSTELGRLQMDQCAKDKKYSL 254
 QY 251 ELGKRAPIVMDADLELAKAIVUSKVINSGVNCABEKVYVQKGIYQFVNRILGEAMQ 310
 255 ELGNAFPIVEEDADILKAVEGALASKFKNAGTQVYANKLYQDGVYDRFAEKLQDAVS 314
 QY 311 AVQGNPAERNDIAMGFLINAAALEREQKVARAVEGAVRAGVAGKGYVYPPPTL 370
 315 KLHIGDGLD-NGYITIGLDEKAVAKVEEHIADALEKARVCGGAKHREGGFTOPTIL 373
 QY 371 LDVROEMSIMHEETFGVLPVAFDLEDAISMANDSYGLTSITYQNLNVMAKAIKL 430
 374 VDVPAANKVSEETFGGLADLPFRKDEADYIAQANDTEFGLAAYFAVARDLSHFVGEAL 433
 QY 431 KFGETIINRNFEMAGCFHAGMKRSKIGGADGKHGILGTLQ 471
 434 EYGIIVGINTGIIISNEVAPFGSIGREGSKGYLEDYLE 474
 Db

RESULT 3
 GARD_RHISN STANDARD: PRT: 491 AA.

AC P55633;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16)
 GN (SSDH).
 GN GARD OR YASJ.
 OS Rhizobium sp. (strain NGR234).
 OC Plasmid sym pNGR234a.
 OC Bacteria: Proteobacteria, alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9719595; PubMed=9164424;
 RA Freiberg C.A., Fellay R., Broughton W.T., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and Legumes";
 RT Nature 387:394-401(1997)
 RL Nature 387:394-401(1997)
 CC -1- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O -
 CC succinate + NAD(P)H.
 CC -1- PATHWAY: 4-aminobutyrate (GABA) degradation
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: A6000096; AAR9149.1; -
 DR HSSP: P05091; ICW3.
 CR 1at1atP3; 1pp020ae; A:b-by3--b-by3r.
 DR Pfam: PF00171; aldehyd. 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR. CYS; FALSE_NEG.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR. GLU; 1.
 KW oxidoreductase; NADP; plasmid.
 FT AC_SITE 263
 FT ACT_SITE 297
 FT SEQ_FEATURE 491 AA; 54253 MW; 6540574401614 cpo64;
 SC

Query Match 33.18; Score 810; DP 1; Length 491;
 Best Local Similarity 47.98; Pred. No. 4; Gaps 2;
 Matches 159; Conservative 85; Mismatches 190; Indels 2; Gaps 2;

26 VNPATFAVISRIPDGGAEEDARKKALDAERKQPEWMLPAIERASMLKISGIERERE 85
 37 VFNSTIGELLAEVPMMAALAHNALEKALAAEPWSGLIAPARSTILKMHFFLEHSD 96
 QY 86 ISATIVFERGKIGGLAEVAVAFYADYIIVYMAFMAFVEDELLQSD 145
 97 LAATITAFEMGKPIGFARKEFVQHAAYIQVAFRANPTVGETISASTDRPMLVIKOPIGV 156
 QY 146 TTGILPMNPFPPLIAPKMAPALLTGNIVIKRSEFTINNAIAFAKTIVPEGLPRGVNLY 205
 157 VCAITPPNFRASNAVARIRISPALAAGCTVVKRPEQTPVAVGAMPALAKIAGPPDVTNLV 216
 QY 206 -LGRGETVGOELACGNPKVAMVSMGTSVSAGFKIMATAKNITKVCLEGRKAPATVMDA 264
 217 VASRPAITSPFTNPVKRISFTGSTFWERILMPQSTQIKRISFLGKNAFPVIFVPA 276
 QY 265 DELAVAKATDSKVINSGVNCABEKVYVQKGIYQFVNRILGEAMQAVGFGAPARNDIA 324
 277 DILAAVVAIAGAFPPNAGTQVYASNPVYASVYAEFAKRTFPPVPTLKVGKSTPPN-VA 335
 QY 325 MGPLINAAALEREQKVARAVEGAVRAGVAGKGYVYPPPTLIDVROEMSIMHEBT 384
 336 IGPLINOPALKKTIHISNADVGARVGGGRRPTSGTFPPPTVTVTSKTMPLAEFT 395
 QY 385 FGVLPPVAVAFDLELAKAIVUSKVINSGVNCABEKVYVQKGIYQFVNRILGEAMQ 444
 396 FGPLAPLILPLDIAHVVREANPTIYSLAAYFASNLKPVAVAFALVYMGVGINQPMSS 455
 QY 445 MGFHAAVMPKSGIGADLKHGILGTLQ 470
 456 EAPFGGVKSGIGREGSRHGLDYL 481
 Db

RESULT 4
 DHAB_BACSU STANDARD: PRT: 490 AA.

AC P71016;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Betaine aldehyde dehydrogenase (EC 1.2.1.8) (BADH).
 GN GBSA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A. AND SPQWNP.F OF 1-25.
 RP STRAIN=168 / J9642;
 RX MEDLINE=96359364; PubMed=8752328;
 RA Koch J., Kempf R., Schmid R., Bremer E.;
 RT "Synthesis of the osmoprotectant glycine betaine in Bacillus
 RT subtilis: characterization of the gbsA genes";
 RT J. Bacteriol. 178:5121-5124(1996)
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;

12b	6	ERK31VLTAKIABELIRDLLEELAELESLDTGKLTLESKDMDDI/ANVQYVYAGLADGDGEI	122
0Y	127	IQSDRGENTNLLFKRALGVTTGTLIPMPNPFLIARKKAVALTGNTIVIKSEPTTNNAI	186
12b	125	ISSDILDSBSKSIILEELIGVQDITPTWNYPLDQASMKIAYALAGNTIVIKKPSRITPLTLTI	184
0Y	187	AFKPIVELEIDEGVEFNIVLDSGETVDFLAGINKVMVSMIGSVASGEIKMATAKNIT	246
12b	185	KVKKIMFPAVAVKQYVANIIVLIRGATVVRPIAVNNIVNIIISPTGIIITPTCKIKMAASGVK	244
0Y	247	KVLEELGKAPAIWMDADLELAKVAIVDSKVINSGVFNCAIRVYVQKGIYDFVNRIG	306
12b	245	KIALELGKKNPIVFKDDELVAVDQALNVAFFHAGVGSAGSKLLVEIDAIHDOFLAELV	304
0Y	307	EAAQAVQFNPAERNDIAMGFLINAAALEREQKVAVAEAGSRVAFKAVEG----	362
12b	305	KKAKPIIKIKIN-GPNAFTSGSPIISAFIRKAVKEYKVEIGLESAKLETGSKPREPELQNG	363
0Y	363	YVYPTPLLLDVROEHSIMHEETGFGVLVVAFTLELVAISMANDSYGLTSSITYQNLNV	422
12b	364	FPEYEPTEFNCNSMDRIYVQFVEFGVITVEFFSSFEPELIELNITVYLACVMSKDIK	423
0Y	423	AMATIKGLKPGCEYINPENFAMGCPHAGWKSITGANDKHGLHGYIOTGVVY	476
12b	424	CPVVAAPLIPMPITVWINPPIYPAQAPWYGRKSGSPFELAKTGLEETVEFHVY	477
RESULT 5			
YQCM_ECOLI			
ID	YQCM_ECOLI	STANDARD:	PRT: 474 AA.
AC	P77674:		
DT	15-JUL-1998 (rel. 36, Created)		
DT	15-JUL-1998 (rel. 36, Last sequence update)		
DT	16-OCT-2001 (rel. 40, Last annotation update)		
DE	Putative betaine aldehyde dehydrogenase (EC 1.2.1.8) (BADH).		
GN	YQCM OR B1444.		
OS	Escherichia coli.		
CC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia		
OX	NCBI_TaxId=562:		
RN	11		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MC1655;		
FX	MEDLINE=9742617; Pubmed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloor G.A., Borna N.T., Burland V.,		
RA	Riley M., Colado-Vides J., Glasner J.D., Kodt C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	*The complete genome sequence of Escherichia coli K-12.*;		
RL	Genome 277:1453-1474(1997).		
RN	121		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		
FX	MEDLINE=9721157; Pubmed=9097039;		
RA	Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,		
RA	Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,		
RA	Kitagawa M., Makioka K., Miki I., Mizubuchi K., Mori H., Mori T.,		
RA	Motomura K., Nakabe S., Nakamura Y., Nishino H., Nishio Y.,		
RA	Osima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,		
RA	Yamano H., Takeda J., Takemori K., Takeuchi Y., Wada C.,		
RA	Yamamoto Y., Horiiuchi T.;		
RT	*A 270-kb DNA sequence of the Escherichia coli K-12 genome		
RT	corresponding to the 28.0-40.1 min region on the linkage map.*;		
RL	DNA RES. 3:363-377(1996).		
CC	-I- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +		
CC	NADH.		
CC	-I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.		
CC	STRONG, TO BADH.		
CC	-----		
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		


```

DB 208 TGNVIVAKTASTSLTSLMAYVSKYIPQAGIPPCVINIVSCKKVGATINNHKIKKAVAT 267
QY 229 GSVANGKRMATANKNTKVCLELGAAPAVIMDDDELAVKAVIVSIVNSGVNCA 488
DB 248 GSTATGRIYOSAAAGKVKVLELGGSPNIVPADLKAAYONILGIIYNSGEVCCAG 327
QY 289 ERVYVOKIYDQFENRIGEMQAVQFENRNDIMAGPLINAAALEREQOKARAVEG 348
DB 428 SRVVEESTYDKFIEEKKASSTKVDIPHESIFGSAATSSMJNKLKLYVDLCKNKG 386
QY 449 ARVAGKAVKRGYVPPILLDVROKSTIMEETPGVLPVAVADTLEDAISMANSD 408
DB 487 ATLTGGERLSSNGYFKPTVPGIVKEDMKIVKEELHGVVIVTKESADEVINMANDE 446
QY 409 YGLTSSIVYTONLANKAKATIGKRGFTYINRPFPMQGFH-----AGAPKSGTGGADCK 463
DB 447 YGLAAGTHHSNINIALKAVDRAVNAAGIYWINFTYN-----DHHAVPFGFNASGLAKEMSV 501
QY 464 HGHAGYLOTOVV 475
DB 502 DALGNTLYOVKAV 513

```

RESULT 7

```

DBAL_AGM1 STANDARD: PRT; 500 AA.
AC 074187:
DE 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (Aldehyde
  ALDH).
GN Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5141:
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=Hort 839;
RA Sheep P.L.; Muller Y.; Vissot J.;
RT "Molecular structure and spatial expression of housekeeping genes in
  mushrooms."
KL Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: Y17825; CAA76875.1;
DB HSSP: P51977; IBS.
DB InterPro: IP002086; Aldehyde_dehydr.
DB Pfam: PF00171; aldhc1.1.
DB PROSITE: PS00070; ALDEHYDE_DEHYDR_YE_1
DB PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU_1.
KW oxidoreductase; NAD.
FT NP_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 269 269 BY SIMILARITY.
FT ACT_SITE 403 403 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54195 MW; CAPCE5HR5085925 Ck064;

```

Query Match

```

Best Local Similarity 45.08; Prod. No. 2.4e 44;
Matches 169; Conservative 104; Mismatches 198; Indels 12; Gaps 9;

```

2 SVVYGHMVIIVGQFVIMKIDAMIVVNPATFAVISTIPGQAFDAKALIAEFA-GPEM 60

```

DB 17 STSINTGLFINGERVDVSKNITLDVVPNPAKGLITTSATRTADIDIAEAHAKAFETW 76
QY 61 EA-LPALEKASWLLKISAGIKERKASELSALIVEBCKIQOLAE-VEVATADYIDYMAEW 118
DB 77 GLNCGSKRGDMYKLAQIMEKNIDLSALEADNCKETPLAKSVSLSTISITHYAGW 136
QY 119 AKRVEGELTOSDRGENIILFKR--ALGVTTGLPNNPFPFLIARMAPALLTGNITVK 176
DB 137 AKKRPVAVETD---EKKLTYSHPPLVAVNQIIPNNPFLIMAKTIGALATNCTYIK 193
QY 177 PSEPTNNALAFATVIRIGIPPGVNIYLGQETVQGFACNPKVAVYSMTSVSAQK 236
DB 194 PSEPTPSALAKMCLALIDEAGFPFGVNVVIVGYSITGQAISSHKIDKVAFTSTLYGK 253
QY 237 IMATAAK-NITKYVLELGKAPAVIMDDNPLAVKAVIVDSVITNSGVNCAFEVYVK 295
DB 254 VMFAAKSNIKNTVLEFGKSPVVTTPDAPTEFSVWTAHGLFNNHQAQAGTPIVQE 313
QY 296 GTPQFVNPVGFAMQAVQFQNPAPNPNTAMGPIINAAALPVPYGVAVAPFGAPAVAVG 355
DB 314 GIVYKPLKKTOKIKELKIDPRTG--IIVGIVSVLDYDRIMSYTESGRAGATVHVGG 372
QY 356 KAVGKGYVPPILLDVROKSTIMEETPGVLPVAVADTLEDAISMANSDVGLTSS 415
DB 373 ERHNGEYFLOPTIFDTTPDMKIVKEELFGPGAVIKERDKEVYIKQANDSNVGLAAV 432
QY 416 YTONLANKAKATIGKRGFTYINREN-PEAQGFHAGMKSSIGSLKKGHAGYLOTOV 474
DB 433 FSDINKAITEIAFAKAGTAWNCANTIDAGVF-GGYKOSIGRHEIYGHANTYVKA 491
QY 475 VYL 477
DB 492 VHV 494

```

RESULT 8

```

DBSSD_RAT STANDARD: PRT; 488 AA.
AC P51650:
DE 01-OCT-1996 (Rel. 34, Created)
DE 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinate semialdehyde dehydrogenase (EC 1.2.1.24) (NAD(+)-dependent
  succinate semialdehyde dehydrogenase).
GN ALDH5A1 OR SSADH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116:
RN 111
RP SEQUENCE FROM N.A. (LINING AND SHORT ISOPRIMS), AND PAPPIAT SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA MDLINB=95113870; Pubmed=7814412;
RA Chambless K L.; Caudle P L.; Hinson P P.; Morgan C F.;
RA Slaughter C A.; Jakobs C.; Gibson K M.;

```

```

FT "Molecular cloning of the mature NAD(+)-dependent succinate
  semialdehyde dehydrogenase from rat and human. cDNA isolation,
  evolutionary homology, and tissue expression."
RL J. Biol. Chem. 270:461-467(1995).
CC -1- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(+) + H(2)O =
  succinate + NADH.
CC -1- PATHWAY: 4-aminobutyrate (GABA) degradation.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
  short form; may be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: BRAIN, PANCREAS, HEART, LIVER, SKELETAL
  MUSCLE, KIDNEY, LOWER IN SPLEEN, JUNG, KIDNEY AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way

```

CC: modified and this statement is not removed. usage by and for commercial
CC: entities requires a license agreement (See <http://www.isib.sib.ch/announcements/>)
CC: or send an email to license@isib.sib.ch

QY 74 KISAGIRERASELSALIVFEGKICQALAEVEVAFTADYIDYMAEWARVEGEFIQSDDPG 133
 Db 126 KWNLMIONKDDARIITAEKSKPLKEANGELIYSAFFLESEFARVYVGIHTPAKO 185
 QY 134 ENILFKRALGVTTGLPNWPFELIARKMAPALLTNGTIVIKPSEFTTNNAIAPAKIVD 193
 Db 186 RRALVLKQPTGVAAVITPNWPFSAITTRKVGAAAGCTVGVVKKPAEDTFFSALALAEAS 245
 QY 194 EIGLPGVENLY---LGRKETVQELAGNPVKVAMVSMTGVSAGAEKIMATAAKNLIKVL 250
 Db 246 QAGIPSGVNVNIPCSKNAKFECEATCTDPLYSKISPTGTTCKILLHHAANSKVRYSM 305
 QY 251 ELGGKAPAVIMDDADLELAKVALVDSRVINSQVCNCAERVVVQKGIYDQFVNRLGAMQ 310
 Db 306 ELGLAPFVDSANVDQAVAGAMASKPRTQTCVCSNQFLVQGIHDAFVKAFEAAMK 365
 QY 311 A-VQGNPARNIDANGPLINAAALFRVPOKVARAVEEPARVAFGGKAVEGGYYPPTL 369
 Db 366 KNLRVNGPFG-FTQGPLINEKAVEKQVNDVASKGATVVTGGKRHQLGKNNFEPTL 424
 QY 370 LLDVAPFMSIMHETEGPVPVVAEDTLEDAISMANDSDYGLTSSIIYQNLNVAMKAIKG 429
 Db 425 LCNVTQMLCTHEETPGPLVAKFEDTEEEAIAANAADVGLAGYFYSQDPAQIWRVAEQ 484
 QY 430 LKFCETVYINRENFEAMQGFHAGWRKSGIGGANGKHKHGLHGYLQTVQV 475
 Db 485 LEVGMVGNELGLSSVECPFGGKQSGLGREGSKYIDLEYLKV 530

RESULT 10

DHA5_YEAST STANDARD: PRT: 519 AA.
 AC P40047;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3).
 GN ALD5 OR ALDH5 OR ALD3 OR YER073W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 NC Saccharomycetales; Saccharomyces; Saccharomycotina; Saccharomycetes;
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97212813; PubMed=9059641;
 RA Wang X., Bai Y., Ni L., Weiner H.;
 RT "Saccharomyces cerevisiae aldehyde dehydrogenases: identification and
 expression.";
 RL Adv. Exp. Med. Biol. 414:277-280(1997)
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dietrich F.S., Mulligan T.T., Hennessey K.M., Allen F., Araujo P.,
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen F., Cherry J.M.,
 RA Chung E., Duncanson M., Guzman F., Hartwell G., Hunkle-Smith S.,
 RA Hyman R., Kayser A., Komp C., Laskari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Ngren P., O'Leary P., Oh C.,
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Rotstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98132377; PubMed=9473035;
 RA Wang X., Mann C.J., Bai Y., Ni L., Weiner H.;
 RT "Molecular cloning, characterization, and potential roles of cytosolic
 and mitochondrial aldehyde dehydrogenases in ethanol metabolism in
 Saccharomyces cerevisiae.";
 RL J. Bacteriol. 180:822-830(1998).
 CC 1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC 2- PATHWAY: Ethanol utilization; second stage.
 CC 3- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
 CC 4- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/homepage/
 or send an email to licen@isb-sib.ch).
 CC EMBL: U56605; AAB01220.1;
 DR EMRL: U18414; AAB44612.1;
 DR HSSP: P51977; IHX5.
 DR SGO: S0000875; ALD5.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldoh; 1
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
 KW oxidoreductase; NAD; Mitochondrion; Transit peptide;
 TRANSIT 1 23 MITOCHONDRION (POTENTIAL);
 FT CHAIN 24 519 ALDEHYDE DEHYDROGENASE;
 FT NP_BIND 265 270 NAD (ADP PART) (BY SIMILARITY);
 FT ACT_SITE 287 287 BY SIMILARITY;
 FT ACT_SITE 321 321 BY SIMILARITY;
 FT CONFLICT 48 48 T -> I (IN REF. 2);
 FT CONFLICT 90 90 L -> A (IN REF. 2);
 FT CONFLICT 93 103 LLEKSVYVPA -> APTKKSIVPE (IN REF. 2);
 FT CONFLICT 410 410 F -> G (IN REF. 2);
 SQ SEQUENCE 519 AA; 56552 MW; 5D333E54477977FC CRC64;
 Query Match 30.7%; Score 750; DB 1; Length 519;
 Best local similarity 33.5%; Pred. No. 4; 6; 4;
 Matches 169; Conservative 102; Mismatches 184; Indels 50; Gaps 9;
 QY 1 MSVPVQHP-----MYIDQFQVTKGDAWIIIVNVATEAVISRIPLQQAEDAKKA 50
 Db 28 LRVPITLPMGFIYEPTGLFTNCEEVASVKQKFTIVINDSNEELIIVYKAMEIDVDEAV 87
 QY 51 DAARAEQPEWALPATERASWIKRISAGIIRRA-----SEISALIVEEGKIQ 98
 Db 88 AALKK-----LLKRSVYCK---AGVRAKALFNLAULVEKHQETLAAIESMUNGKSL 135
 QY 99 ELAEVFAVFAIYIYMAFWAFKPEFETLQDQENILIF KALGVITIGLFWNPF 156
 Db 136 FCARGIVAVLSYLSQGGWAKIYGNVITD---GRNHTYSIKREPLAVVQIIPWNPFL 192
 QY 157 FLIAEKMAFALLIINTIVIKSEETTRNAIAIAIVDEELIGVIRLVIQGGTYSDEL 216
 Db 194 LMSWKIGIPALATGNTVVKIETTPLSALCASLQDENGIPAVVNIIPSGPVVQV 252
 QY 217 ANPFFVAMVSMLESVAPEFIMATAAKNLIIPWTLIIGKALAIIVMGAGAFVAKAIVDS 276
 Db 253 SAIDIVKRIAPTGTSTATGHIIMKVAATVVRKVTIIEGKSNIVIAIADGLVAKNIAPG 312
 QY 277 RVINSQVCNVAEPVVVAPFIYIYFVNRLGAMQAVQGNFAEKNDIAMSHLIANAALER 336
 Db 313 IFYNCEVCCASRIYIQVTVVEVLQKIVYTESLKWDDPTVEVCACTSKVQLHR 371
 QY 347 VEQVAPAVEPARGVAFGKAVEGGYYPPTLLIDVHQEMSIMHEETFGVLPVAVPT 396
 Db 372 ILDVEVAKSEGARLVTSAGHIGSKGYFVKVTVFAVKMEIMFVKEEVFGFIVTSKFSI 431
 QY 397 LEDAISMANDSIYGLTSSIIYQNLNVAMKAIKLGKPGCTYINRGNFAMQGFH 451
 Db 432 VDEIVAMANDSQGLAAGIHTNDINKVAVSKVKAKATVWNIYN NFRNVVTFG 486
 QY 452 WRKSGIGGADKIKHIGYLOTVQV 475
 Db 487 FGSGIGREMGFAALSNVITQKSV 510
 RESULT 11
 DHAM_LEITA
 ID DHAM_LEITA
 AC Q25417; PRT: 498 AA.

204 ALYFASLCKKVGIVPAGVNNIVIPGIPVIGAAIINIIPIPIKIAPITISTIVKSAUVSSS 263

QY 244 NIKVKLEGGKAFVAIVMGDAELEAVKAIVSRVINSVGVN'AEPVYGIVGIDQEVN 403
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
264 NLKKITILEGCKSAHLVFICANTKTPTNVINCIEPKNAQTICSSGPSIVGEGIVDELLA 323

QY 304 FLGRAMIA-VUFGNPFAFPNIANGPLINAAALHPVBLKVAFARAKVAFIRKAVRDK 462
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 AFKAYLETEIKVNPPFIKARF-QSALINQQEFTIMNYILDGKKESAKITITGFKVRKNG 382

QY 363 YYYPFTILLIVRGEMSMHETETSPVLVPVATTELEPAISMANISNGTTSSTVTQNUNV 422

Db 393 YFIRPTFYDNDMRIVKEIFGPVVTVAKTKTLLEGVEMANSSEFGLOGSIGTESLSLT 442

QY 423 AMKAIKGLKGQEIYINPEN-FFAMCGFHAGSPKSGTGADGRKHGHGYLOTGVYLQ 478
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 443 GLKVAKMLKACTVMINTYDNDFDSRVPF-GSVKQSGVGRPMGEVHYAYTEVKAVPIK 498

RESULT 15

DHAB_SPIOL STANDARD; PRT; 497 AA.

ID AC PI7202; P93535;
DM 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)

DE betaine aldehyde dehydrogenase, chloroplast precursor; (EC 1.4.1.8) (BAHD).

OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
OX NCBI_TaxId=3562;
RN [1]

SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

PP STRAIN=cv Savoy hybrid 612;
PC MEDLINE=96207274; PubMed=2320587;
RA Weretilnyk F.A.; Hanson A.D.;
RT "Molecular cloning of a plant betaine-aldehyde dehydrogenase, an
ET enzyme implicated in adaptation to salinity and drought.";
KL Proc. Natl. Acad. Sci. U.S.A. 87:2745-2749(1990).
RL [2]

RN SEQUENCE FROM N.A.
RP Shu W., Ai W., Chen S.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -! CATALYTIC ACTIVITY: Betaine aldehyde + NAD(-) + H(2)O = betaine +
NADH

CC -! PATHWAY: Betaine biosynthesis; last step.

CC -! SUBUNIT: HOMODIMER.

CC -! SUBCELLULAR LOCATION: Chloroplast.

CC -! SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

This SWISS Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.ebi.ac.uk/staff/ebl/ebdoc/0000/> or send an email to license@ebi-sib.ch.

EMBL: M31480; AAA34025.1; -
DE EMBL: 063412; AAB41636.1; -
PIR: A35944; A35944
DR HSP: P05091; ICM3.
DE InterPro: IPRO02086; Aldehyde_dehydr.
LP Pfam: PF00171; aldhdh; 1.
LP PROSITE: PS00070; ALDEHYDE_DEHYDR_CVS; 1.
LP Kofas1f: KS0687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Chloroplast; Transit peptide.

FT TRANSIT 1 7 CHLOROPLAST (POTENTIAL).
FT CHAIN 8 497 BETAIN-ALDEHYDE DEHYDROGENASE.
FT MCP_PES 8 497 BLOCKED.
FT NE_BIND 235 240 NAD (ADP PART) (BY SIMILARITY)

```
FT ACT_SITE 257 257 BY SIMILARITY
FT ACT_SITE 291 291 BY SIMILARITY
FT CONFLICT 424 424 S -> F (IN REF. 2)
SQ SEQUENCE 497 AA: 54270 MW: 55088240663522F CRC64;

Query Match
Best Local Similarity 34.2%; Score 718.5; Dh 1; Length 497;
Matches 162; Conservative 104; Mismatches 203; Indels 19; Gaps 8;

QY 1 MSVPV-CHPMYIDGQFVTRGDA--WIDVNPATEAVTSRIPDGOAEDARKAIDAAPRA 56
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 MAEPFAPARQLFIDGE---WREPTKKNRPVNPSTEEIIGDIPAATAEDVEVAVVAARRA 57

QY 57 --OPEWALPAITERASWLRKISAGIRERASEISALIVERGGKIQOOLAEVEVAFTADYIDY 114
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 58 FRNNWSATSCAHRATYLRALIAAKITEKKDHFVKLETTIDSGRPDEAVLDDDDVASCFEY 117

QY 115 MAEWARKYEGE-----IISDRPGENILFKRALGVTTGILPNFPFFELIARKMAPALLT 169
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 118 FAGQAFALDCKQKAPVTLPMERFKSHVL--RQPLGVVGLISPNYPLLMATWKIAPALAA 175

QY 170 GNTIVIKPSEFTTNNALIAFAKIVDEIGLPRGVFNVLGRGETVQQLAGNPKVAMVMTG 229
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 176 GCTAVLKPSLAVTCLERGEVCNEVGLPPGVNLTGLGPDAGAPIVSHPDVDKIAFTG 235

QY 230 SVSAGEKIMATAAKNITKVCLLELGGKAPAIVMDDADLELAVKAIIVDSRVINSGOVCNAE 289
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 236 SSATGSKVMASAAQLVKPVTLELGGKSPVVFEDVDIDKVPVPTIFGGEFWNGQICSAT 295

QY 290 KVVYQKGYDQVYNRIGEAQVQFONPAERNL IAMGPLINAAALERVEQKVARAVEEGA 349
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 296 RLLVHESIAAEFVDKLVKWKTKNIKISDPPEEG-CRLGPVLSKGQYDKIMKFISTAKSRA 354

QY 350 RVAFGGKAVE--CGKGYYPPTLLLOVROEMSIMHEETGCVLPVVAFTDLEDAISMANDS 407
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 355 TILYGGSRPHLKKGYIEPTIVTDISTSMQIMKEEVFGPVLCKVTESSEDEAIALANDT 414

QY 408 DYGITTSIYQNLNVAMKAIKGLKFGETYINRENFEAMOGFHAGWRKSGIGGADGKHGLH 467
  | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 415 EYGLAAVAFSNDLCERITKALEVGAVVWNC:SQPCFVQAPWGGIKRSCFGRELGEWGIG 474

QY 468 GYLQTOVV 475
  | : : : : : |
Db 475 NYLNIKGV 482
```

Search completed: June 24, 2003, 10:14:29
Job time : 10.6475 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:08:15 ; Search time: 46.000 seconds

(without alignments)
2862168 Million col: updates/sec

Title: us-09-830-751-6

Perfect score: 2445

Sequence: 1 MSVPVQHPYIDGQVPTWKG ADEKHGHSYVQGVVVYLS 474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phase.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvrius.*
- 17: sp_bacteriap.*
- 18: sp_cheap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2426	99.2	474	16 Q8X901	Q8X901 Escherichia
2	1557	63.7	480	16 Q8X904	Q8X904 Neisseria m
3	1552	63.5	480	16 Q9JW97	Q9JW97 Neisseria m
4	1279.5	52.3	393	16 Q9P119	Q9P119 Campylobact
5	1066.5	43.6	486	2 Q93088	Q93088 Microscilla
6	903	36.9	493	17 Q9H001	Q9H001 Thermoplasma
7	902	36.9	493	17 Q97900	Q97900 Thermoplasma
8	894	36.5	492	16 Q9PKF1	Q9PKF1 Streptomyces
9	879.5	36.0	483	16 Q9T6M5	Q9T6M5 Pseudomonas
10	872.5	35.7	497	16 Q9Y3P2	Q9Y3P2 Ralstonia s
11	870.5	35.6	482	16 Q8X950	Q8X950 Escherichia
12	868.5	35.5	481	16 Q92MM2	Q92MM2 Salmonella
13	858.5	35.1	492	16 Q824F8	Q824F8 Salmonella
14	845.5	34.6	475	16 Q9K7P5	Q9K7P5 Bacillus ha
15	840.5	34.4	484	16 Q921F2	Q921F2 Rhizobium m
16	836	34.2	491	16 Q92V65	Q92V65 Rhizobium m

17	835	34.1	468	16 Q9KE63	Q9KE63 bacillus ha
18	832	34.0	486	16 Q9CKS0	Q9CKS0 Pasteurella
19	831.5	34.0	485	16 Q9ML54	Q9ML54 rhizobium l
20	827	33.8	528	16 Q9SEK4	Q9SEK4 Arabidopsis
21	822	33.6	482	16 Q9A3R2	Q9A3R2 Enterobacter
22	821.5	33.6	487	16 Q9Y157	Q9Y157 Brucella m
23	814.5	33.3	484	16 Q92Y70	Q92Y70 rhizobium m
24	814.5	33.3	509	16 Q9SAK4	Q9SAK4 Arabidopsis
25	813.5	33.3	485	16 Q98711	Q98711 rhizobium l
26	811.5	33.2	488	16 Q9YK14	Q9YK14 listeria mo
27	809	33.1	488	16 Q92DA3	Q92DA3 listeria in
28	804	32.9	495	16 Q9R277	Q9R277 deiscoccus
29	803	32.8	482	16 Q94438	Q94438 bacillus su
30	801.5	32.8	474	2 Q948F1	Q948F1 rhodococcus
31	798.5	32.6	486	16 Q8U4H1	Q8U4H1 agrobacteri
32	797.5	32.6	478	17 Q82V31	Q82V31 pyrobaculum
33	797	32.6	509	5 Q9VBP6	Q9VBP6 drosophila
34	796.5	32.6	484	2 Q9VPD7	Q9VPD7 agrobacteri
35	795.5	32.5	489	16 Q8UKH2	Q8UKH2 agrobacteri
36	795.5	32.5	490	16 Q922L8	Q922L8 rhizobium m
37	795	32.5	494	16 Q92X00	Q92X00 rhizobium m
38	794	32.5	497	16 Q98AY0	Q98AY0 rhizobium l
39	793.5	32.4	493	16 Q8XPW7	Q8XPW7 raietonia s
40	791.5	32.4	487	16 Q9W346	Q9W346 rhizobium m
41	785.5	32.1	484	2 Q9WMD5	Q9WMD5 agrobacteri
42	784.5	32.1	484	16 Q8U852	Q8U852 agrobacteri
43	783.5	32.0	553	16 Q91R16	Q91R16 oryza sativ
44	782.5	32.0	474	16 Q8X965	Q8X965 escherichia
45	781.5	32.0	477	16 Q9JTN7	Q9JTN7 neisseria m

ALIGNMENTS

RESULT 1

Q8X901	Q8X901	PEPELINAPY:	PRT:	479 AA.
AC	Q8X901:			
DT	01-MAR-2002 (TREMBLrel_20, Created)			
DT	01-MAR-2002 (TREMBLrel_20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel_21, Last annotation update)			
DE	Aldehyde dehydrogenase, NAD-linked.			
GN	ALDH A22306 OR ECS-021.			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=83334;			
RN	[1]			
SP	SEQUENCE FROM N.A.			
ST	STRAIN=0157:H7 / F10333 / ATCC 700927;			
KX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.T., Mayhew G.F., Evans P.S., Gregor T., Kirkpatrick H.A.,			
RA	Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Weich R.A., Blattner F.R.;			
RA	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";			
PL	Nature 409:529-533(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
FX	STRAIN=0157:H7 / K1M0.0509952;			
FX	MEDLINE=2114231; PubMed=11254796;			
PA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
PA	Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
PA	Jida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
PA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
ET	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
ET	O157:H7 and genome comparison with a laboratory strain K12";			
FL	UNA Res. 8:11-22(2001).			
FL	EMBL: AA005964; BAB35444.1;			
DE	EMBL: AA005964; BAB35444.1;			
DE	Interpro: IIR02006; Aldehyde_dchidr.			

```

DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 479 AA; 52238 MW; D49819FA02E5C7CF CRC64,

Query Match
Best Local Similarity 99.2%; Score 2426; DB 16; Length 479;
Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSVPQVPMYIDGQFVTRGDAWLDVYNPATEAVISKIPGQAEADAKKALDAERAQPEW 60
DB 1 MSVPQVPMYIDGQFVTRGDAWLDVYNPATEAVISKIPGQAEADAKKALDAERAQPEW 60
QY 61 EALPAIERASWLKISAGIRERASEISALIVEEGGKIQQALAEVEVAFTADYIDYMAEWAR 120
DB 61 EALPAIERASWLKISAGIRERASEISALIVEEGGKIQQALAEVEVAFTADYIDYMAEWAR 120
QY 121 RYEGEITQSDRPGENILLFKRALGVTTGILPWNPFPFLIARKMAPALLTGNTIVIKPSEF 180
DB 121 RYEGEITQSDRPGENILLFKRALGVTTGILPWNPFPFLIARKMAPALLTGNTIVIKPSEF 180
QY 181 TTNNAIAFAKIVDEIGILPRGVNVLGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 240
DB 181 TTNNAIAFAKIVDEIGILPRGVNVLGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 240
QY 241 AAKNITKVCLELGGKAPAVMDADLELAVKAIIVDSRVNSGVCNCAEKVVYQGIYDO 300
DB 241 AAKNITKVCLELGGKAPAVMDADLELAVKAIIVDSRVNSGVCNCAEKVVYQGIYDO 300
QY 301 FVNRLGAMQAVQGNPAERNDTAMGPLINAAALERVEQKVAKEGAVAFGKAVG 360
DB 301 FVNRLGAMQAVQGNPAERNDTAMGPLINAAALERVEQKVAKEGAVAFGKAVG 360
QY 361 KGYTYPPTLLDVRQEMSIMHEETFGPVPVAFDTLEDAISMANSDVGLTSSIVQNL 420
DB 361 KGYTYPPTLLDVRQEMSIMHEETFGPVPVAFDTLEDAISMANSDVGLTSSIVQNL 420
QY 421 NVAMKAIKGKFGETYINRENFEAMQGFHAGWPKSGIGGADKGHGLHYLOTVQVYVLS 479
DB 421 NVAMKAIKGKFGETYINRENFEAMQGFHAGWPKSGIGGADKGHGLHYLOTVQVYVLS 479

RESULT 2
Q9JW97
ID Q9JW97 PRELIMINARY; PRT; 480 AA.
AC Q9JW97;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Aldehyde dehydrogenase A.
GN NMB1966.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickoy E.K.,
RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark F.R.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.K., Rappaport R., Winter J.C.;
PT *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.*;
RL Science 287:1809-1815(2000).
DR EMBL: AE002544; AAP42297.1;
DR HSSP: P51977; 1BXS.
DR TIGR: NMH1966; 1.

```

```

DR PROSITE: PS000676, ALDEHYDE_DEHYDR_GLY; UNKN-WN_1;
DR PROSITE: PS000687, ALDEHYDE_DEHYDR_GLU; UNKNOWN_1;
KW Oxidoreductase, Complete Proteomes
SQ SEQUENCE: 480 AA; 52355 MW; 1504807076347609 CRC64;

Query Match
Best Local Similarity 63.6%, Score 1552, DB 16; Length 480;
Matches 300; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHPMYIDGQFVTRGDWIDVNVNATPAVTSRIPDQGAEDARKAIDAERAQPEWALPA 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3 QIANYINGRENFENGERVNYNFEETAEIAREPKGKADVDRAVAARAQAQAPWRLPA 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 66 IERASWLRKISAGIPEPASEYTSALIVBGGKLCGLAEEVAVFTADYIDYMAEWARVEGE 125
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 63 VEPGAYLPTAGIPEPADELDTYIVAPGCKTKPLAPVEVFTADYLYQAEWARVEGE 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 126 ITOSDRPRENILLKRAIGVTTGILPNWPFELIARKMAPALLTGTNTIVIKPSEFTTNA 185
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 ITQSDPPENITLLEKPTDVGAGILPNWPFELIARKMAPALLTGTNTIVIKPSSVTTFIN 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 186 IAPAKIVDEIGLPGCVENLVLRGETVGOELAGNPKVAMVSMTGSVSAGEKIMATAKNI 245
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 183 HIFAEIVDAVGLPAGVENVNPGCAEIGNALSAHPQVMVSLTGSVEAGSPQVWEAASANI 242
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 246 TKVCELGSKAPATVMDGADLELAVKAIIVISKVINSGLVGNCAEKPVYVVGK5YDQFVNR 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 243 TKVSELGSKAPATVMDGADLELAVKAIIVISKVINSGLVGNCAEKPVYVVGK5YDQFVNR 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 306 GEMAOVQGNP--AERNDIAMGLINNAALERVQKVARVEGAVGKRAVEGGY 363
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 303 TAAKGVRYGNPAEAGALEMPLIEERAVKAVKVERAVKOGAKLVCGGGRABRGY 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 364 YYPPTLLLDVROEMSGIMHEETFGVPLVPAFDTLEDALSMANDYGLTSSIVYTONLVA 423
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 363 FEPTLLDTDSMDIMKEETFGVPLVPAFDTLEDALSMANDYGLTSSIVYTONLVA 422
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 424 MFAIKGLKFGETYINRENEFAMQPHAGWKSSTGTAAGKHLGYSGLTAVVYLQS 479
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 423 FYVTRFLGFGETYINRENEFAMQPHAGWKSSTGTAAGKHLGYSGLTAVVYLQS 478
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 4
Q9P119 PRELIMINARY: PRT: 393 AA.
AC Q9P119;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Aldehyde dehydrogenase C-terminus (EC 1.2.1.22).
GN ALD' OR CJO490.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria, epsilon subdivision, Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168.
PX MEDLINE=26150412; PubMed=1688264;
RA Parkhill J., Wren B.W., Mungall K., Ketley T.M., Churcher C.,
RA Basham D., Chillingworth T., Davies K.M., Fellwell T., Holtroyd S.,
RA Tagels K., Karlyshev A.V., Moule S., Pallao M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
PT reveals hyper-variable sequences."
PL Nature 403:665-668(2000).
DR EMBL: AL139675; CAB75128.1;
DR HSSP: P51977; 1BXS.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aided; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKN-WN_1.
KW Complete proteome.

SQ SEQUENCE: 485 AA; 42907 MW; 1504711000F4E6AA7 P-64;

Query Match
Best Local Similarity 63.1%, Score 1279.5, DB 16; Length 393;
Matches 245; Conservative 52; Mismatches 96; Indels 1; Gaps 1;

QY 91 VEEQKIKGLAEEVAVFTADYIDYMAEWARVEGEYFIIDSDPPGENILLKRAIGVTTGIL 150
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 MOEQKTRVLAETIENFTADYMDYTAEWARYEGEIIOSDRANEHIYLYKSAIGVIGGIL 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 151 PNWPFELIARKMAPALLTGTNTIVIKPSEFTTNAIAFAKIVDEIGLPGCVENLVILGPG 210
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 PNWPFELIARKMAPALLTGTNTIVIKPSEFTTNAIAFAKIVDEIGLPGCVENLVILGPG 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 211 TVGCELAGNPKVAMVSMTGSVSAGEKIMATAKNIITKVCLEGGKAPATVMDGADLELAV 270
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 121 VVGVELSSENIIMVSLTGSVEAGTKVNEAAKNLIKVSLELGGKAPATVMDGADLELAV 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 271 KATVSPVINSQVNTAFRVYVVGKIGYDGFVNHLAFAMGAVAFGNPAEPNDIAMGPLIN 330
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 EALKASKRICNNQVNTAFRAYVHTSVYDREVKFKVAMSKVSGNTI-KGDFUMGPLVN 239
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 331 AAALPVPVQKVAQVAFVAFVAFVAFVAFVAFVAFVAFVAFVAFVAFVAFVAFVAFV 390
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 240 QAVVUNALALQKATAGALVEGSKLIDTSYEPASVLINVKHETIMUKLEAFAPILP 259
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 391 VVAFVDEIGLCAISMANISAGELTSSIVYTONLVAAMKAIKGLKHLGYSGLTAVVYLQ 450
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 300 LAKPDTLDEIVIMANDYGLTSSIVYTONLVIAMKASPEIKFQETYNKRNPRAMQSPHA 459
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 451 GWPKSGTGGAGKHLGYSGLTAVVYLQ 478
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 360 GFRKSGIGGAGKHLGYSGLTAVVYLQ 387
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 5
Q93P88 PRELIMINARY: PPT: 485 AA.
AC Q93P88;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE MS143, putative aldehyde dehydrogenase.
OS Microscilla sp. PREL.
OC Bacteria; CFB group; Flexibacter group; Microscilla.
OX NCBI_TaxID=155537;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PREL.
RA Zhong Z., Toudarian A., Helinski D., Knauf V., Sykes S.,
RA Wilkinson J.F., O'Ryne C., Shea T., De Loughery C., Caspi R.;
RT "Sequence Analysis of a 101 kb Plasmid Required for Agar Degradation
of a Microscilla strain."
RI Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
CC "SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DE EMBL: AF339846; AAK67865.1; -.
DE InterPro: IPR000067; Aldehyde_dehydr.
DE Pfam: PF00171; aided; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Oxidoreductase; Plasmid.
SQ SEQUENCE: 485 AA; 53309 MW; 94FFELU3Q2EQAAGGK P-64;

Query Match
Best Local Similarity 43.6%, Score 1066.5; DB 2; Length 486;
Matches 208; Conservative 100; Mismatches 161; Indels 21; Gaps 4;

QY 1 MSVPVQHPMYIDGQFVTRGDWIDVNVNATPAVTSRIPDQGAEDARKAIDAERAQPEWALPA 52
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3 MSVKNYKLYIDGE-----WIDATGESQELSPIDTETVGVTVMGVGAQALEA 54
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 53 AEPAPPEWALPAEPASWLRKISAGIPEPASEYTSALIVBGGKLCGLAEEVAVFTADY 112
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

Db 55 AERAKQKVPARKRADLRTFAAEIKANKNOLAELIVREOGKLLAVAKMEVEVTASFI 114
 QY 113 DYMAEWARRYEGEIIQSDRGENTILLPKRALGVTTGILPWNFFFLIAKKMAPALLTQNT 172
 Db 115 EYACDGAKSIEGDIIPSDNPAEHIIMHKIPRGVVVAITAWNFFPLAAGKLGPAIVAGNS 174
 QY 173 IVIKPSETTNNALAFKIVDEIGLPRGVNVLVRCGTGVGOELACNPKVAMVSMTGWS 232
 Db 175 VVLKPTQPTPLATLEIGNANKVGLJPKGLNIIITGOGSLGNALVAMPITKVVSMTGSP 234
 QY 233 ACEKIMATAAKNITKVCLELGGKAPALVMDADLELAVKAVDSRVINSQVCNCALRVY 292
 Db 235 AGGGLFRTAENLEHVLELGGKAPCIVFDADLEAVEGAFHSRFDNCGVCTSNERY 294
 QY 293 VOKGIYDQFVNRKLGCAQVQGNPAPRNDIANGPPIINAAALERVEQKVARAVEGARVA 352
 Db 295 VHEISYNEFFERFMEKVKYGLKLGNDPMDA-ATTIGPKVNAKEVAHMEHLVTKSVEGATVA 353
 QY 353 FGKRAVEG- ---KGYYPPPTLLLDVROEMSIHHEETFGVPLVPVAFDTLEDALSMANDSD 408
 Db 354 IGGKKPOCKAFAKCHWFEPITLTDVKONMTIVHEESFGPLPVKIFESEFNEVIGYANDCE 413
 QY 409 YGLTSSYTONLVAMKAIKGLKFGETYINRENFEAMQGFHAGWRKSGIGGAGKHHJHS 468
 Db 414 YGLAAVFTDMNKMIMRINDELEFGEIYINRGHGEHQHGFHNGYKLSGTGGEDSKYGFEO 473
 QY 469 YLQTVVVYLO 478
 Db 474 YMEKKTFFIK 483

RESULT 6

Q9HK01 PRELIMINARY; PRT; 493 AA.
 AC Q9HK01: 2001 (TrEMBLrel. 16, created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Probable aldehyde dehydrogenase.
 GN TA0809.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatiales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2403;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed:11079001;
 RA Ruegg A., Graml W., Santos-Martinez M.-I., Koroike Y.F., Volker G.,
 RA Newes H.-W., Frishman D., Stocker S., Lupas A.N., Baumelster W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513(2000).
 DR EMBL: AL445065; CAC11938.1;
 DR ISSP: P05091; ICW3
 DR InterPro: IPR02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd. 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 493 AA; 54782 MW; 2F6A79DD96BE7FF5 CRC64;

Query Match 46.9%; Score 903; DB 17; Length 493;
 Best local Similarity 48.4%; Pred. No. 9,3e-54;
 Matches 182; Conservative 100; Mismatches 184; Indels 8; Gaps 3;

QY 9 MYIDQGVTVWRGDWIDVVPATEAVTSRIPDQAEADARKADAAPRAQPEWALPATER 68
 Db 5 LYIDQGVWVSSNGDITKYNPVTGVEVIGKEAASRDIDRAIDADAFQKMMIGSVSR 64
 QY 69 ASWLKISAGIRERASEISALIVEEGRKIQQLAEVEVAFATAYIDYMAEWARKYEGEITTO 128
 Db 65 SKLIYRAKELIENKRAELNIIEENGKPKVKEAEEVDGVIDQIYYAEAWARKLNGEVVE 124
 QY 129 SDKPGENILLKRALGVTTGILPWNFFFLIAKKMAPALLTQNTIVIKISEETTNNALAF 188
 Db 125 GTTNIRKIFQYKVPYGVTVVVALTPWNPFGAMVARKLAPALLTGNITVILKISSDTPGSAWI 184

QY 129 SDRPGENILLKRALGVTTGILPWNFFFLIAKKMAPALLTGNITVILKISEETTNNALAF 188
 Db 125 GTTNIRKIFQYKVPYGVTVVVALTPWNPFGAMVARKLAPALLTGNITVILKISSDTPGSAWI 184
 QY 189 AKIVHEICELDPRGVENIVDEHGETWQGLA-INKVAMVSMTGWSA-ERIMATAAKNIFV 248
 Db 185 VPKEVEAGVPKGVNLITGHCSEFSGYIVPHKKVNLITMGSTATGQPIQMKASANKAKI 244
 QY 249 CLELGGKAPALVMDADLELAVKAVDSRVINSQVCNCALRVY 292
 Db 245 ILELGGKAPVWVKDALIMENALKTLLWAKYWNAGQSP-IAAEPIYVHEIYIDTFMSKEVEL 404
 QY 409 MOAVQFQNPAPENDIANGPPIINAAALFPVFGKVARAVEFEAPVAPK AVES; KUY 463
 Db 305 SRKIALGTPKNAP- MPTGINKGALQAT-PIVEEAKESGAK-ITGDSQDSLSGYPHNY 461
 QY 364 YYPPTLLDVRQEMSIHHEETFGVPLVPVAFDTLEDALSMANDSD 408
 Db 362 FELPTIIGNADQSKIFUREIFAVV-APP-SSVFTMYI-ANDSKYGLASYLEFTKIPNII 421
 QY 424 MKAIKGLKFGETYINRENFEAMQGFHAGWRKSGIGGAGKHHJHS 468
 Db 422 FEASFRTPFGETVYNNMFGPASPQYHTEPMTGQAG-PSKYGISTPYIKIKNIYV 475
 RESULT 7
 Q979W0 PRELIMINARY; PRT; 493 AA.
 AC Q979W0:
 DT 01-OCT-2001 (TrEMBLrel. 18, created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Aldehyde dehydrogenase.
 GN TVI050 OR TVG1076124.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatiales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=50339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GSSI / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed:11121011;
 RA Kawashima T., Anano N., Koike H., Makino S. I., Hiiuchi S.,
 RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanohori K., Kawamoto T.,
 RA Nunohiiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
 RT "Archaeal adaptation to higher temperatures revealed by genome
 sequence of Thermoplasma volcanium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14757-14762(2000).
 DR EMBL: AP000995; BAB60192.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd. 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 493 AA; 54801 MW; EDEYGERH292TJ54 CRC64;

Query Match 36.9%; Score 902; DB 17; Length 493;
 Best local Similarity 38.4%; Pred. No. 1.1e-54;
 Matches 182; Conservative 96; Mismatches 188; Indels 8; Gaps 2;

QY 9 MYIDQGVTVWRGDWIDVVPATEAVTSRIPDQAEADARKADAAPRAQPEWALPATER 68
 Db 5 LYIDQGVWVSSNGDITKYNPVTGVEVIGKEAASRDIDRAIDADAFQKMMIGSVSR 64
 QY 69 ASWLKISAGIRERASEISALIVEEGRKIQQLAEVEVAFATAYIDYMAEWARKYEGEITTO 128
 Db 65 SKLIYRAKELIENKRAELNIIEENGKPKVKEAEEVIVILQMLYYAFWARKLNGEVVE 124
 QY 129 SDKPGENILLKRALGVTTGILPWNFFFLIAKKMAPALLTQNTIVIKISEETTNNALAF 188
 Db 125 GTTNIRKIFQYKVPYGVTVVVALTPWNPFGAMVARKLAPALLTGNITVILKISSDTPGSAWI 184

RT *Complete genome sequence of enterohemorrhagic *Escherichia coli*
PT Q157:H7 and genomic comparison with a laboratory strain K-12.*

BL DNA Pos 8-11-22(2001)
DR EMBL: AE005495; AAC57768.1
DR EMBL: AP002562; BAB36945.1
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh.1
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 482 AA; 51848 MW; 00F30470AFAE45 CPO64;

Query Match 35.6%; Score 870.5; DB 16; Length 482;
Best Local Similarity 39.3%; Pred. No. 2.1e-51;
Matches 181; Conservative 92; Mismatches 187; Indels 1; Gaps 1;

QY 11 IDQFVTRQDAWIVVNPATRAVTSPTDGGAGELAPKRAITAAEPAQEWALVAIKPAS 70
DB 15 INGEWLDANNGEVIDVNFANRKRKSSVFKMGADPTAAIIAANFALPFWALIAKFFAN 74
QY 71 WLKKISAGIPEKASELSALIVERGGKIGGLAPVEVAFADXLIVYMAEWAPYVEGEIIQSD 130
DB 75 IIRNWNLMHEHQUOLAKMLLEQKKPLAEAKGEISYAASFLWFAEKGKLYIDTPGSH 144
QY 131 PPGENTILFKPALGVITGIIIPNFFFLIAPKMAPALLIGNIIVIKPSEETINNAIAFAK 190
DB 135 QADKRLIVIKQPIGVTAATIPWNPFAAMITRKAGPALAAGGIMVLKIPASQIFPSALALAE 194
QY 191 IVDEIGLPRVNLVLCRQETVQGLACNPKPKYAMVMTGSSVSAEKKIATATAKNIKVL 250
DB 195 LAIRAGIPAGVFNVTGSAGVGNELTSPNPKRISPTGSTEISGLMEQCAKDIKKVSL 254
QY 251 ELGGKAPAIWDDADLELAKAIVDSRVINSQVNCNCAERVVQKGIYDQVNRIGPAMQ 310
DB 255 ELGSAFFIVEGDAULDKAVEGALASKFENASQTCVANKLYVGVVDFRPAFKLQAVS 314
QY 311 AVQFGNPAERNDIAMGPLINAAALEPVEGKVAFAVEEGARVAFGGKAVGGYYPPTLL 370
DB 315 KLVIGDLQKG-VTIGLIDKRAKAVEKHIAALEKGAHVVGKGAHERGSENFQETIL 373
QY 371 LDVROEMSIMHEFTFGPVLVPAVDFTLEDAISMANDSYGITSSTYTGNLVANKAIKGL 430
DB 374 VDFPANAKYSKEETFGPLAFKLEALVIAANUTFEGLAAYFYAKULSKVFRVSEAL 444
QY 431 KFEETYNENFAMQGFHAGWKSGIGGADCKHGLHYDQ 471
DB 434 EYGIVGINTGIITSNEVAPFGIKASGLGKSGSKYGIETYDLE 474

RESULT 12

Q8ZMM2 EPELIMINAPY. PRT: 482 AA.
AC Q8ZMM2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Succinate-semialdehyde dehydrogenase 1, NADP-dependent (EC 1.2.1.16).
GN GABD OR STM2791.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
[1]
RN SEQUENCE FROM N.A.
RP STPAIN-LT2 / SCS03412 / A100 760720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sandersson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Powellik S., Ali T., Danta M., Du Rou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of *Salmonella enterica* serovar Typhimurium

PI LT2.*

EL Nature 413:852-856(2001).
DP EMBL: A000627; AAC21676.1;
DR InterPro: IPR002086; Aldehyde_dehydr
DR Pfam: PF00171; aldedh.1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 482 AA; 51848 MW; C58E8F508D35097D CPO64;

Query Match 35.5%; Score 868.5; DB 16; Length 482;
Best Local Similarity 39.4%; Pred. No. 2.1e-51;
Matches 182; Conservative 91; Mismatches 188; Indels 1; Gaps 1;

QY 10 YLDGQFVWKGSIAMIVVNPATRAVTSPTDGGAGELAPKRAITAAEPAQEWALVAIKPAS 69
DB 14 FIDHFWPFAAPGSDIVIPVSNPANPKLISNVPKMGCAETPDAINANPALPAMPALTAKERA 73
QY 70 SWIPKISAGIPEKASELSALIVERGGKIGGLAPVEVAFADXLIVYMAEWAPYVEGEIIQSD 129
DB 74 NILKFWNLMMHEHQUOLAKMLLEQKKPLAEAKGEISYAASFLWFAEKGKLYIDTPGSH 133
QY 140 DHPGENTILFKPALGVITGIIIPNFFFLIAPKMAPALLIGNIIVIKPSEETINNAIAFAK 189
DB 134 HQADKRLIVIKQPIGVTAATIPWNPFAAMITRKAGPALAAGGIMVLKIPASQIFPSALALA 193
QY 190 KIVDEIGLPRVNLVLCRQETVQGLACNPKPKYAMVMTGSSVSAEKKIATATAKNIKVC 249
DB 194 ELAEPANITPAQVFNVTGSAGVGNELTSPNPKRISPTGSTEISGLMEQCAKDIKKVS 253
QY 250 LELGGKAPAIWDDADLELAKAIVDSRVINSQVNCNCAERVVQKGIYDQVNRIGPAMQ 309
DB 254 LELGSAFFIVEGDAULDKAVEGALASKFENAGQTCVANKLYVGVVDFRPAFKLQAV 313
QY 310 QAVQFGNPAERNDIAMGPLINAAALEPVEGKVAFAVEEGARVAFGGKAVGGYYPPTLL 369
DB 314 NKLAVDGLQGA-DVAGLPLIDKRAKAVEKHIAALEKGAHVVGKGAHERGSENFQETI 372
QY 370 LLDVROEMSIMHEFTFGPVLVPAVDFTLEDAISMANDSYGITSSTYTGNLVANKAIKGL 429
DB 373 LADVPDPAKYSKEETFGPLAFKLEALVIAANUTFEGLAAYFYAPDLSRPFVGEA 432
QY 440 KFEETYNENFAMQGFHAGWKSGIGGADCKHGLHYDQ 471
DB 433 LEYSIGVINTGIITSNEVAPFGIKASGLGKSGSKYGIETYDLE 474

RESULT 13

Q8ZAF8 PRELIMINARY. PRT: 482 AA.
AC Q8ZAF8
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Succinate-semialdehyde dehydrogenase (EC 1.2.1.16).
GN SLI2911.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OC Salmonella.
OX NCBI_TaxID=601;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.P., Pinkard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.R., Sebahin M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies P.M., Dowd L., White N., Farrar J.,
RA Patwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
PA Krogh A., Larsen T.S., Leather S., Meule S., O'Gaora P., Pary C.,
PA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
PA Whitehead S., Barrrell B.G.;
RT *Complete genome sequence of a multiple drug resistant *Salmonella*

RT enterica serovar Typhi CT18.":
RL Nature 413:848-852(2001).
DR EMBL: AL627276; CAD05900.1; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 482 AA; 51894 MW; FF850C52F818DC7 CRC64;

Query Match 35.1%; Score 858.5; DB 16; Length 482;
Best Local Similarity 39.2%; Pred. No. 1e-50;
Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

QY 10 YDGFVFWKGDAMVDVNVNPAFAVSKIPUGDAEAPKAIADAAFAQPEWPAIPATERA 69
DB 14 FIDGWRDARGSDVIPSVPANGKPIGNVPKMAETETRDINAANPAIPAWPAITAKERA 73
QY 70 SWLRKISAGIRERASEISALIVEGGKIQOLAEVEVAFTADYIDYMAEWARRYEGEITQS 129
DB 74 NTLRWFLMMEHODDLARLMTLEQKPLAEAKGEISYAAASELEWFAEBCRIYGDITPC 133
QY 130 DRPGCNILLFKRALGVTTGILPNWPFELIARKMAPALLIGNTIVIKPSEFTTNNAL 189
DB 134 ROTDKRLVVKQPIGVTAALTPMNPSPAMITRKAGPALAAGCTMWLKPASQTPFSALALA 193
QY 190 KIVDKIGLPGVFNVLVGGTGVQELAGNPKVAMVSMTGSVSACEKIMATAAKNITKVC 249
DB 194 ELAQRAQILAGVFNVTGSGAGTGGELTSNPIVPKLSFTGSTEIGPQLMECAKDKKYS 253
QY 250 LELEGKAPAIYMDADILELAVKAIIVDSRVINSQVNCARVYVVGKIVDQFVNRISEAM 309
DB 254 LELEGNAFIVFDADILKAVEGALASKFPNAGQTVCANPIYVQGVYDRFAEKLQAV 313
QY 310 QAVQFGNPAERNDIAMGPLINAAALERVEOKVARVEEGARVAFGCKAVEGKYYPPTL 369
DB 314 NKLAVGDLQA-DVAIGPLIDERAKAVQEHADALEKRGARVITGEAHLGNFQPTI 372
QY 370 LLDVQKQMSIMHETFGPVIIPVAFDTLEDAISMANDSYGLTSSITYTONLNVAMKAIKG 429
DB 373 LADVPDAKVAKEFTGFLAPLFRFSDADVIRQANUTEFLGAAYFYARDLSRVFVGEA 432
QY 430 LKFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLO 471
DB 433 LEYGVIGINTGIIISNEVAPRGCIKASGLSPGSKYGLIENYLE 474

RESULT 14
Q9K7P5
ID Q9K7P5 PRELIMINARY; PRT; 475 AA.
AC Q9K7P5.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DE Succinate-semialdehyde dehydrogenase.
GN GABD OR BH3316.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058112;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki P., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001518; BA807035.1; -.
DR HSSP: P51977; 1BX5.
DR InterPro: IPR002086; Aldehyde_dehydr.

DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 475 AA; 51447 MW; 9A1A2E8E8F9C117 CRC64;

Query Match 44.6%; Score 845.5; DB 16; Length 475;
Best Local Similarity 39.1%; Pred. No. 7.7e-50;
Matches 184; Conservative 84; Mismatches 195; Indels 7; Gaps 4;

QY 9 MYIDGQPVTKGDA--WIDVNVNPAFAVSKIPUGDAEAPKAIADAAFAQPEWPAIPATERA 66
DB 6 LYIDG---TWIGDLDLTFVKNPATGEVGVGMINGKAEAAAAIEAQRAFIKWKYIAA 62
QY 67 ERASWLKRISAGIRERASEISALIVEGGKIQOLAEVEVAFTADYIDYMAEWARRYEGEIT 126
DB 53 EKANFLMTWHLQILAQKFFIAPMLIMPKGCLAFAPFETYSASFIPWLAFFETPVVYGT 122
QY 127 IOSDRPGCNILLFKRALGVTTGILPNWPFELIARKMAPALLIGNTIVIKPSEFTTNNAL 186
DB 124 VVASAEAKPITHVQKQV-VVAAALTPMNPFAAMEVPKMAIALAACTFVCKPAFIPTAV 182
QY 187 AFAKIVDQFGNPAERNDIAMGPLINAAALERVEOKVARVEEGARVAFGCKAVEGKYYP 246
DB 183 KIVELTEPAFPKGVVWLVVGGAKFTGPEMSEPHVVKITFTGSTAVSKIMKQSAIRM 242
QY 247 KVCLEGGKAPAIYMDADILELAVKAIIVDSRVINSQVNCARVYVVGKIVYQFVNRI 306
DB 243 NESTFIDGAPMILITGADILELAVKAVVAVKVPVNAQGVVKNVLYVGLVVEKPKPT 302
QY 307 EAMQAVQFGNPAERNDIAMGPLINAAALERVEOKVARVEEGARVAFGCKAVEGKYYP 365
DB 303 SKVVQVLEGGKLEES-VHIDGELTEKRYKAKRAIVDAVAKARVYVVGKQGNISY 361
QY 366 PPTLLDVRQKMSIMHETFGPVIIPVAFDTLEDAISMANDSYGLTSSITYTONLNVAMK 425
DB 362 LPTLTUVHLEMLVMQFETETPVAFIQLALETGVLEKANGIDYLAAYFIITKAPLIP 421
QY 426 AIKGLKFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLOTVV 475
DB 422 LSPALDFGVVWNCAPSTACADPFGMKPFGSGPGRSQGFAPLETFEV 471

RESULT 15
Q92TE2
ID Q92TE2 PRELIMINARY; PRT; 484 AA.
AC Q92TE2.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Probable succinate-semialdehyde dehydrogenase (EC 1.2.1.16).
GN GABD OR K00014 OR SMC02780.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN 1021;
RX MEDLINE=21396507; PubMed=11481440;
RA Capella D., Barloy-Hubler F., Gouzy J., Boile G., Ampe F., Barut J.,
RA Gollstard P., Becker A., Boutry M., Cadieu F., Dreyer S., Gloux S.,
RA Goudier T., Goffeau A., Kahn D., Kiss P., Leloutre V., Masny D.,
RA Pohl T., Portetelle D., Puchler A., Porcelle B., Ransperger U.,
RA Renard C., Theault P., Vandenbol M., Weidner S., Galliard F.;
RT "Analysis of the chromosome sequence of the legume symbiont
Rhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591782; CAC01401.1; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.

SEQUENCE:	484 AA:	51866 MW:	AAC2988AAZ9D5D498	CRC64;
Query Match:	34.4%:	Score 840.5:	DB 16:	Length 484;
Best Local Similarity	37 %:	Pred No 1	7a-4q:	
Matches 174:	Conservative	97:	Mismatches 186:	Indels 9: Gaps 2;
QY	20	GDAW-----LDVVNPATEAVISKIPUGOAEJARKAIDAAEAKQAPEWEALPAIFRASW	71	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	16	GETWFADPKNAIEVNNPATGTEIGFVPKLGAAETPTALEAAARVKQEWAAARTAKERSAV	75	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	72	LKTSAGISIRPASEISEAIIVFEFGGIQLAQEAVEVAFTADYTDYMAENAPPYEEIITQSDR	131	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	76	LRRWFELMIENKKDLGRILTMFGKPLAEATGEIVYGASFIEWFAEARRYGLDVPGHQ	135	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	132	PGENILLFKAALGVTTTGILPNWPFFFLAPKMAPAILTGNITVIKPSFFTINNAIAFAKI	191	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	136	KDKRILVNKPQIGVVAATIIWNFPNMITTKAGPALAACAMWLKPPAAQTFFSAIAJVL	195	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	192	VDEITLPRGVFNILVLRGETVGVOELAGNPKYAMVSMTGSVSAGEKIMATAAKNTIKVCLE	251	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	196	AEFAGMPKGLFSVITSAPETCAEMTSNTPVKLTFTGSTEVGAELVQSNATIKKLGLE	255	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	252	LGGKAPAIVMDDADLELAIKVAIVDSKVINSQOVNCABERYVYVKGITYDFQVNRIGEAMQA	311	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	256	LGSNAPTIVFTTHAFLLAAVFESAIJAEPHNGGTGCANKLYVQGVYEATSDKLAQAVAK	315	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	312	VQCNPAERNDIAMGPLINAAALERVCKVAHAVEEGARVAFGCKAVEGKGYYPPTLLL	371	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	316	LKTNGME-DGVTLGLPIDOPALKVKEEHVADALAKRVAVOGGRRHSLSGTFYEATVLA	374	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	372	DVRFNMSTMHFFETFGPVLPVAFDTLEDIAISMANSDYGLTSSITVTONLVANKAIKGLK	431	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	375	DVTOAMAVAREETFGVPAPLFRPKQESDVIQAQANTEFGLASYFYAKDLARVFRVAELE	434	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	432	PGETYINPENFAMQCETHAGNPKSGIGGADGKHGLHYLOTVQVYL	477	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		
QY	435	YGMVGVNGLISTAEAPFGCVKSLIGREGSKYGEEMEIKYYCCL	480	
Ddb		: : : : : : : : : : : : : : : : : : : : : : : : :		

Search completed: June 24, 2003, 10:17:13
Job time : 40 6606 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:25 ; Search time 40.3175 Seconds
(without alignments)
1696 383 million cell reads

Title: US-09-830-751-8
 Perfect score: 2673
 Sequence: 1 MTNNPSSAOIKPGEYGFPLK...HYQCTKCLVSYSDKPLNLF 512

Scoring table: PROSUM62

scoring table	ProSim2	Gapop 10.0	Gapext 0.5
---------------	---------	------------	------------

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 slides

Database

A_Geneset1_lnl002 *	
1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT **	
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT **	
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT **	
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT **	
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT **	
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT **	
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT **	
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT **	
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT **	
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT **	
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT **	
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT **	
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT **	
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT **	
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT **	
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT **	
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT **	
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT **	
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT **	
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT **	
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT **	
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT **	
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT **	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2673	100.0	512	22	AA874926	E. coli aldehyde d	
2	1714.5	64.1	506	22	AA972418	C glutamicum prote	
3	1714.5	64.1	506	22	AA879351	Corynebacterium gl	
4	1682	62.9	506	22	AA849208	V cholerae vpi pha	
5	1505.5	56.3	1302	22	ABG35490	Novel human diagno	
6	1505.5	56.3	1393	22	ABG32965	Novel human diagno	
7	1186.5	44.4	496	22	AA034957	Enterococcus faeca	
8	1090.5	40.8	508	18	AA889793	Staphylococcus aur	
9	1080	40.4	510	23	ABP19646	Staphylococcus epi	
10	1074	40.2	371	22	ABG27689	Novel human diagno	

ALIGNMENTS

RESULT 1	
AA874926	
TF	AA874926 standard; protein; 512 AA.
XX	
XX	
XX	AA874926;
XX	
XX	26-JUN-2001 (first entry)
XX	
XX	E coli; a,3-hydroxy-dehyd-4r-glycose; aldP protein sequence; EF2 ID NO:8.
XX	
XX	Alb-hydro-dehyd-ogenase; glycerol dehydratase; 3 HP; glycerol;
XX	feedstock; 3 hydroxypropionic acid; genetic engineering; glucose;
XX	bacterial host; absorbable prosthetic device; surgical suture;
XX	beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
XX	polyhydroxalkonate; copolymer; lactic acid.

PT 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 PT enzymes in the presence of glycerol or glucose -

XX Claim 5; Page 46-48; 63pp; English.

CC The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (I); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhap gene from *Klebsiella pneumoniae* and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (I). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid
 CC incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the *E. coli* aldehyde dehydrogenase aldH,
 CC which is used in the exemplification of the present invention.

XX Sequence 512 AA;

Query Match 100.0%; Score 2673; DB 22; Length 512;
 Best Local Similarity 100.0%; Pred. No. 1.3e-264;
 Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 MTNPPSAQIKPGEGYFPLKARYDNFVIGGEWVAPADGEYVYVNTPTVGTQLLCEVASSG 60
 DB 1 MTNPPSAQIKPGEGYFPLKARYDNFVIGGEWVAPADGEYVYVNTPTVGTQLLCEVASSG 60
 QY 61 KRDLIDALDAHKVKWKAITSVQDRAALFKIADRMENONLELLATAETWNGKPIRTS 120
 DB 61 KRDLIDALDAHKVKWKAITSVQDRAALFKIADRMENONLELLATAETWNGKPIRTS 120
 QY 121 AADVPLAIDHFRFASCIHQEGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 DB 121 AADVPLAIDHFRFASCIHQEGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 QY 181 MAPALAGNCVVLKPARLTPISVLLMEIVGDLIPGCVVNVNAGGVTCGYLATSKRTA 240
 DB 181 MAPALAGNCVVLKPARLTPISVLLMEIVGDLIPGCVVNVNAGGVTCGYLATSKRTA 240
 QY 241 KVAFTGSTEVCQIMQVATONIPVTILEGGRSPNIVFAVMDDEDAFFDKALEGFALFA 300
 DB 241 KVAFTGSTEVCQIMQVATONIPVTILEGGRSPNIVFAVMDDEDAFFDKALEGFALFA 300
 QY 301 FNOEVCVTCPSRALVQESIVERPMERAIIPVVESTPSNPLDSVTOMGAGVSHQLETLIN 360
 DB 301 FNOEVCVTCPSRALVQESIVERPMERAIIPVVESTPSNPLDSVTOMGAGVSHQLETLIN 360
 QY 361 YIDIGKKEGALVITGSPKKILEGELKGYLEPTILFPGNNMKVFEELIPGLVAVTFK 420
 DB 361 YIDIGKKEGALVITGSPKKILEGELKGYLEPTILFPGNNMKVFEELIPGLVAVTFK 420
 QY 421 TMEALELANDTQYGLGAGVWSNGLATKMGRTQAGRVWNTCYHATPAHAAGFGYKQS 480
 DB 421 TMEALELANDTQYGLGAGVWSNGLATKMGRTQAGRVWNTCYHATPAHAAGFGYKQS 480
 QY 481 GIGRETHKMKI FHYQOTKTTLVSYSDKPLGLF 512
 DB 481 GIGRETHKMKI FHYQOTKTTLVSYSDKPLGLF 512

RESULT 2
 AAC92818
 ID AAC92818 standard; protein: 506 AA.
 XX

AC

AAC92818;

XX

26-SEP-2001 (first entry)

XX

C glutamicum protein fragment SEQ ID NO: 6572.

XX

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW

organic acid synthesis.

XX

Corynebacterium glutamicum.

OS

EP1108790-A2.

XX

20-JUN-2001.

XX

18-DEC-2000; 2000EP-0127688.

XX

16-DEC-1999; 99JP-0377484.

PR

07-APR-2000; 2000JP-0159162.

PR

03-AUG-2000; 2000JP-0280988.

XX

(KYOWA) KYOWA HAKKO KOGYO KK.

PA

Nakagawa S, Mitozuchi H, Ando S, Hayashi M, Ichiai K, Yokoi H;

PI

Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

WPI: 2001 376931/40.

DR

N-PSDB: AAH68037.

XX

Novel polynucleotides derived from Coryneform bacteria, for identifying
 mutation point of a gene, measuring expression of a gene, analysing
 expression profile or pattern of a gene and identifying homologous gene

PS

Claim 17; SEQ ID NO: 6572; 24pp + Sequence Listing; English.

XX

The present invention provides a number of nucleotide and protein
 sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
 are useful for identifying the mutation point of a gene derived from a
 mutant of coryneform bacterium, measuring expression amount and
 analysing the expression profile or expression pattern of a gene derived
 from Coryneform bacterium, and identifying a homologue of a gene derived
 from coryneform bacterium. Coryneform bacteria are useful for producing
 amino acids, nucleic acids, vitamins, saccharides and organic acids,
 particularly L-lysine. The present sequence is a protein described
 in the exemplification of the invention.

CC

Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the
 European Patent office.

XX

Sequence 506 AA;

SQ

Query Match 64.1%; Score 1714.5; DB 22; Length 506;

Best Local Similarity 62.5%; Pred. No. 2.2e-166;

Matches 314; Conservative 82; Mismatches 105; Indels 1; Gaps 13

QY

12 PGYGFPLKARYDNFVIGGEWVAPADGEYVYVNTPTVGTQLLCEVASSGKRDIIDALAA 71

DB

7 PGEGSVNVEKRYENFGCKWVYFVGVYVNTPTVGTQLLCEVASSGKRDIIDALAA 66

QY

72 HKYDKWHAISVQGRALIFKIAKMEONLELLATAETWNGKPIRTSAAVPLAIDH 141

DB

67 HAAADWGRKTSVAERALIIRIADRMEEHEERIVAVETWENKAVRETLAALPLAIDH 126

QY

132 RYFASCTRAEGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWKMAFALAA 191

DB

127 RYFAGAI RAEGRSSQIDHNTVAYHFHEPLGVVGQIIPWNPFLLMATWKALALAA 186

QY

192 VLKPAKLEPLSVLLMEIVGDLIPGCVVNVNAGGVTCGYLATSKRTAKVAF 251

DB

187 VMKPAQVHASTLLYLNIIGDLIPGCVVNVNAGGVTCGYLATSKRTAKVAF 246

QY

252 QQIMQVATONIPVTILEGGRSPNIVFAVMDDEDAFFDKALEGFALFA 411


```

DB 247 KLINRAASUKIIEVILELSEKSPSEIFSIW: SQUFAFAKAVESCFAMFALNGEVCICYS 309
QY 312 RALVQSEIYEFMEFAIRFVESIESNLLSVLWMAJAVSHGULETILNYIDIGKKEGAD 371
DB 307 RALVHESIADFELELVKRVQNIKLGNDLDTMTMGAGASQOMDKISSYKIGPPEGAG 366
QY 372 VLTGGKRLKLEGLKQYYLEPTILLFGUNNMHVEEIEFGPVLAVITFKIMEALELAND 431
DB 367 TLGGKVNKVDG-MENGYIETPTVFGTNMWFPPFETFGVLSVATFSDFDEAIRIAND 425
QY 432 TQVGLGAGVMSRNGNLAYKMRIGTQAGRWVWVNYHAYFAHAAPGSGYKSGIGRETHKML 491
DB 426 TNYGLGAGVMSRQNTIYRAGRAIOAGRWVWVNYHAYFAHAAPGSGYKSGIGRETHKML 485
QY 492 EHYQOTKCLLVSDKPLGLF 512
DB 486 NHYQOTKCLLVSDKPLGLF 506

RESULT 3
AAB79351
ID AAB79351 standard; Protein; 506 AA.
AC AAB79351;
XX 30-APR-2001 (first entry)
XX Corynebacterium glutamicum SMP protein sequence. SEQ ID NO 278.
XX Corynebacterium glutamicum. Sugar metabolism and energy production.
KW SMP protein, sugar metabolism and oxidative phosphorylation protein,
KW fine chemical production, organic acid, proteinogetic amino acid,
KW nonproteinogetic amino acid, purine base, pyrimidine base, nucleoside,
KW nucleotide, lipid, saturated fatty acid, unsaturated fatty acid, diol,
KW carbohydrate, aromatic compound, vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX W0200100844 A2.
XX 04-JAN-2001.
XX 23-JUN-2000; 2800W0-1500943.
XX 25-JUN-1999; 990S-0141031.
PR 08-JUL-1999; 990E-1031412.
PR 08-JUL-1999; 990E-1031413.
PR 08-JUL-1999; 990E-1031414.
PR 08-JUL-1999; 990E-1031420.
PR 08-JUL-1999; 990E-1031424.
PR 08-JUL-1999; 990E-1031428.
PR 08-JUL-1999; 990E-1031431.
PR 08-JUL-1999; 990E-1031433.
PR 08-JUL-1999; 990E-1031434.
PR 08-JUL-1999; 990E-1031510.
PR 08-JUL-1999; 990E-1031562.
PR 08-JUL-1999; 990E-1031634.
PR 08-JUL-1999; 990E-1032180.
PR 08-JUL-1999; 990E-1042227.
PR 08-JUL-1999; 990E-1042230.
PR 08-JUL-1999; 990E-1043208.
PR 14-JUL-1999; 990E-1032924.
PR 14-JUL-1999; 990E-1032973.
PR 14-JUL-1999; 990E-1033005.
PR 27-AUG-1999; 990E-1040765.
PR 31-AUG-1999; 990S-0151572.
PR 03-SEP-1999; 990E-1042076.
PR 04-SEP-1999; 990E-1042086.
PR 04-SEP-1999; 990E-1042087.
PR 03-SEP-1999; 990E-1042088.

```

```

PR 04-SEP-1999; 990E-1042095.
PR 03-SEP-1999; 990E-1042123.
PR 03-SEP-1999; 990E-1042125.
XX (BAD) BASF AG.
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
DR WPI; 2001-061975/07.
DR N-PSDB; AAF71468.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes.
XX
PS Claim 20; Page 450-452; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogetic
CC or nonproteinogetic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (II) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein sequences required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 506 AA;
Query Match 64.1%; Score 1714.5; DB 22; Length 506;
Best local Similarity 62.5%; Pred. No. 2,2e-166;
Matches 313; Conservative 82; Mismatches 105; Indels 1; Gaps 1;
QY 12 PGEYGFPLKAPYDNFVIGGEWVADGYYONLTPVTGQLCEVASSGKXRDIDLDA 71
DB 7 PTEGSTVNVYKPYENYIGKRWVPPVVEGYLENISPVTCEVCFVARGTAHVELA 66
QY 72 HKYKDKKRAHTSVQDPAALIFKLTADMEQNLLELAFWNGCKPIPETSAADVLAI 131
DB 67 HAAADAWGKTSVAEPALILHPTADPMEEHLEIAVETWENTZAVPETLAADIFLAI 126
QY 132 RYFASGIRAOEGGISEVDSETVAYHFHEPLGVVGGIIPWNPFLLMASWKPALAA 191
DB 127 PYPAGATPAQDFPSSQIDHNTVAYHFNEPIGVVGGIIPWNPFLMATWKLAPALA 186
QY 192 VLKPAKLLPLSVLLMEIVGDLPLPGVWVWVWAGAGVIGEYLATSKRTAKVATG 251
DB 187 VMKFAEGGIIASLYLINIISGLIPEISVLNVSDNGTEAQAALSSNFRKTAFT 246
QY 252 QQIMOVATIGNITPVTLELSEKSFNIVADVMEESVAFDFALDEALFAENGGEV 311
DB 247 KLINFAASUKLIIVILELSEKSPSEIFSIW: SQUFAFAKAVESCFAMFALNGE 306
QY 312 RALVQSEIYEFMEFAIRFVESIESNLLSVLWMAJAVSHGULETILNYIDIGKKEG 371
DB 307 RALVHESIADFELELVKRVQNIKLGNDLDTMTMGAGASQOMDKISSYKIGP 366
QY 372 VLTGGKRLKLEGLKQYYLEPTILLFGUNNMHVEEIEFGPVLAVITFKIMEAL 431
DB 367 TLGGKVNKVDG-MENGYIETPTVFGTNMWFPPFETFGVLSVATFSDFDEAI 425
QY 432 TQVGLGAGVMSRNGNLAYKMRIGTQAGRWVWVNYHAYFAHAAPGSGYKSGIGRE 491

```


XX 21-MAR-2001; 2001W010500180
 PF 21-MAR-2001; 2000US-1910780
 PR 23-MAY-2000; 2000US-206848P
 PR 26-MAY-2000; 2000US-207727P
 PR 23-MAY-2000; 2000US-242578P
 PR 27-NOV-2000; 2000US-253625P
 PR 22-DEC-2000; 2000US-257931P
 PR 16-FEB-2001; 2001US-269308P
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS52816.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 10550; 511bp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 496 AA:
 SQ
 Query Match 44.4%; Score 1186.5; DB 22; Length 496;
 Best Local Similarity 48.6%; Pred. No. 2.9e-112;
 Matches 238; Conservative 79; Mismatches 164; Indels 9; Gaps 4;
 QY 25 YDNFIFGHWAPANGAEEYQNIITVVGQILCEVASSCKRFDIDLDAHKVKYKWAHTSVQ 84
 DB 14 YQLYINGKWTTCGSKNKMIASYNPSNGEKLAEFFVDATNADVDRAVEAAQFAQTWKDIDVV 73
 QY 85 DRAALFLKIAURMEONLELLATARTWNGKPIRETSAADVPLAIDHFRYPFASCIQAQEGG 144
 DB 74 TRSNLLIKLADLIENOEHLAMWETLNGKPKRETSTGVPSADHFRYPFASVIRGFEZG 133
 QY 145 ISEVUSITVAYHFRPGVGVGQIIPWNPFLIMASWKMALPAAGNCVWLKPARLTPLSVL 204
 DB 134 VKFEKDTLSIVKREPCVGVGQIIPWNPFLIMGAWKLAPALAAAGNTVTHPSSSTLSLL 193
 QY 205 LIMEIVGDLINCVNWNVNGAGGVIGIYLSATSKRIAKVAFSTGVGGQIMVATQNIIP 264
 DB 194 ELKIFIQVLPKGVNLIITGRGSDSGNYMLAHPGFDKLAFTSTGVGYTVAKAAADRIIP 253
 QY 265 VTIELGKSPNIVPADVMDERDAFFOKAIFGFAL-FAFNGEVCCTCPSPALVQESLYEPF 323
 DB 254 ATLEGGKSNIPF.....EDANWEPALEGVOLGILFNGQVCCAGSRVFGSGIYGF 307
 QY 324 MERAIRRVESIRSNPLSDSVTQMGAAQVSHQIETILNYIDIGKKGADVLTGGRKRLLEG 383
 DB 308 VEALKERFEQVNGFPWEKQVEMGAQINHQLEILKYVEIGVKEGATLITGGQR-LTEN 366

QY 384 ELKDYTYLLEFLLF GANNMEVIVQELIICVLAVYIIKIMLEALLLARGCYGLGAVWS 442
 DB 367 GIDKGAFLAPTLLANETINIEVAQFFIFGIVAVIKFEETEEVIRIANISEYGLGAVFS 426
 QY 443 KGNLAVKMGKGLGAGLVWNTNYHAYPAHAAPGCGYKSGICGRTTHKMMLEHYGGTCGLV 502
 DB 427 GQIRVALEVAEIVPTGPMWVNYHLLACALISAKSGISFELHSSMLIAYTGMKNYI 466
 QY 503 SYSDKPIGLF 512
 DB 487 VTKEADGLY 496
 XX
 XX RESULT 8
 XX AAM89794
 XX ID AAM89794 standard; Protein: 508 AA.
 XX AC AAM89794;
 XX XX
 XX 1b-MAR-1999 (first entry)
 XX
 XX Staphylococcus aureus protein Seq ID #5241.
 XX
 XX Computer readable medium; vaccine; *S. aureus* infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome.
 XX
 XX Staphylococcus aureus.
 XX
 XX EP/86519-A2.
 XX PN 30-JUL-1997.
 XX
 XX 07-JAN-1997; 47FP-0100117.
 XX
 XX 05-JAN-1996; 96US-0009861.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Choi GH, Dillon PJ, Fannon MK, Kunsch CA;
 PI Rosen CA;
 XX
 XX WPI: 1997 374922/45.
 XX
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 PT stored on computer readable medium and used in the production of
 PT anti-*S. aureus* vaccines
 XX
 XX Claim 23; Page 3242-3244; 4271pp; English.
 XX
 CC This sequence represents a *Staphylococcus aureus* protein sequence of the
 CC invention. The DNA sequences encoding the *S. aureus* proteins are recorded
 CC on a computer readable medium, preferably selected from a floppy or hard
 CC disk, random access memory (RAM), read only memory (ROM) or CD ROM.
 CC Homology searches using the *S. aureus* DNA sequences allows putative
 CC functions to be assigned so that protein encoding or regulatory regions
 CC of commercial, therapeutic or industrial importance can be obtained.
 CC Specifically, sequences which are likely to encode antigens have been
 CC identified and these polypeptides can be used in a vaccine composition
 CC against *S. aureus* infection. The polypeptides can also be used in a kit
 CC for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated
 CC in numerous human diseases, including cellulitis, eyelid infections, food
 CC poisoning, osteomyelitis, skin and surgical wound infections, scalded
 CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the
 CC DNA sequences can be used for recombinant production of the polypeptides.
 CC The new DNA sequences (and their fragments) are useful as primers on
 CC probes for isolating homologues of any of the 514 *S. aureus* DNA sequences
 XX contained on the computer readable medium.
 XX
 XX Sequence 508 AA:

XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 24-AUG-2000; 2000US-064167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS91876.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 58048; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. Amino acid sequences AR330377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 371 AA:
 Query Match 40 28; Score 1074; DR 22; Length 471.
 Best Local Similarity 90 58; Pred No. 5, 9c-101;
 Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 301 FNGCEVCTCPSRALVQESIVFERMERAIRRVESIRSGNPLISVTOMCAQVSHGQLETLN 360
 DB 165 FNGCFVCTCPSPALVQESIVFERMERAIRRVESIRSGNPLISVTOMCAQVSHGQLETLN 224
 QY 361 YIDICKKKGADVITCGPRKILLEGKDYILEPTILPGQNNMVFQEEIFGCVLAVTTPK 420
 DB 225 YIDICKKKGADVITGPRPKILLEGKDYILEPTILPGQNNMVFQEEIFGCVLAVTTPK 484
 QY 421 TMEAELELANDTQYGLGAGVWSRNGNLAYKMGKRGIOAGRWNTNCHYHAYPAHAAGFYGYKOS 480
 DB 285 TMEAELELANDTQYGLGAGVWSRNGNLAYKMGKRGIOAGRWNTNCHYHAYPAHAAGFYGYKOS 344
 QY 481 GIGRETHKMMLLEHYQOQTKCLLVSYSD 506
 DB 345 GIGRETHKMMLLEHYQOQTKCLLVSYSD 370
 RESULT 11
 AAB74924
 ID AAB74924 standard: Protein: 500 AA.
 XX

AC AAB74924;
 XX 26-JUN-2001 (first entry)
 DT Human aldehyde dehydrogenase ALDH2 protein sequence SEQ ID NO:4.
 DE
 XX
 XX Aldehyde dehydrogenase; glycerol dehydrogenase; 4 HP; glycerol;
 KW feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;
 KW bacterial host; absorbable prosthetic device; surgical suture;
 KW beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
 KW polyhydroxyalkonate; copolymer; lactic acid.
 XX
 OS Homo sapiens.
 XX
 XX WO200116346-A1.
 PN
 XX 08-MAR-2001.
 PD
 XX 30-AUG-2000; 2000WO-US24878.
 PF
 XX 30-AUG-1999; 99US-0151440.
 PR
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Suthers PF, Cameron DC;
 XX WPI: 2001-315988/33.
 DR N-PSDB; AAF82082.
 XX
 XX 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 PT enzymes in the presence of glycerol or glucose
 XX
 PS Claim 5; Page 32-35; 63pp; English.
 XX
 CC The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises forming a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism (i) expresses genes for non native
 CC enzymes which catalyse the production of (3-HP) from (I); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydrogenase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3 HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhap gene from Klebsiella pneumoniae and a gene for
 CC an ADH capable of catalysing the production of (3 HP) from (I). 3 HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or for incorporation into beta-lactams,
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the best organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3 HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the human aldehyde dehydrogenase ALDH2, which
 CC is used in the exemplification of the present invention.
 XX
 SQ Sequence 500 AA:
 Query Match 45 78; Score 954; DR 22; Length 500;
 Best Local Similarity 42 08; Pred. No. 2, 4c 88;
 Matches 206; Conservative 84; Mismatches 176; Indels 24; Gaps 10;
 QY 28 FFGCFWVAPAGFYQYQNI TIVTQJLCEVASSKRPDPLAD- AAIKVKIKWALISVQ 84
 DB 24 FFINEMWDVSRKTEPTPTNPTSTGEVTCQVAESDKELVDKAREGRPGAFQAGSPWRHMVAS 84
 QY 85 DRAALFKIALRMHONLELLATAETWNGKPIRETSAALVPLADIDHRYFAS-TPAQEKS 144
 DB 84 HSGELNPLALAEKPEKFTYLALELLNRPYVISTLVGLDVLNLEPYPAWALRYDCK 144
 QY 145 LSEVDSPTVAYHFIEPLGVVQDII PNWFTLLMASWMAFALAAAGN-VVLKPAEGLEPSVL 204
 DB 144 TPIDGGFESYTRIEHPGVVQDII PNWFTLLMASWMAFALAAAGN-VVLKPAEGLEPSVL 204

```

QY 205 :LMEIVGSL-IPPGVWVWVNGAGVIGEYLATSKETAKVAFTGSTVEGQIMQYA-IQNI 262
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 204 YVANLIKEAGPPGVVNIIVPGFGPTAGAAIASHEDYDKVAFTGSTEIGRIVQVAAAGSNL 263
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 IPVTILELGGKSPNIVFADVMDEDAFEDKALE GEALFAENGVEVCTCPSEALVQESLY 320
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 KRVTILELGGKSPNIIINS-----DAUMUNAVEAHEALF-FNUGUCCAGSKIPVQEDLY 316
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 321 ERFMERAIRVESTRSNPFLDSVTQMGAVSHQGLEIILNYITIGKKKGAVLVTGSRPKL 380
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 317 DEFVVSVAFAKSVVGNPNFTSKTEGSPQVDEIFKFIILGYINTQKQGAKLCCG---- 372
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 381 LEZELKD--GYYLEPTILFG--QNNMKVFOREIFGPVLAVTTFKTEAEALELANDTQYIGL 437
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 --GIAADRGVFTQPTV-FRIVQDGMITAKFEIFGPMQILKFKTIEEVVGPANNSTYGLA 429
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 438 AGVSRPNQNIAYKMGRTIQAGPVNTWYHAYPAHAAFGAYKQSGISGTFPHKKMMLHEHYQT 497
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 430 AAVFTKDLDRANYLISQALQAGIYVWNCYIVFGQSPFGYKMGSGRGREILGEYGLQAYTEV 489
   . . . : ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 498 KCLLVSYSDK 507
   . . . : | | | | |
Db 490 KTVTKVPQK 499
   . . . : | | | | |

RESULT 12
ABB60140
ID: ABB60140 standard: Protein: 520 AA.
XX
AC ABB60140:
XX
DT 26-MAR-2002 (first entry)

```

[illegible]

Search completed: June 24, 2003, 10:13:47
Job time : 42.2175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: Jun 24, 2003, 10:13:55 Search time: 14 1793 seconds
(without alignments)
1062 435 Million cell updates/sec

Title: US-09-830-751-8

Perfect score: 2673

Sequence: 1 MTNPPSAQINPGEVGFPLK HYDQKLLVSYSHKPLDPLF 512

Scoring table: ELCSUM62

Gap 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6R_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1080	40.4	510	4	US-09-134-001C-4541
2	867.5	32.1	521	4	US-09-221-294-2
3	847	31.7	485	4	US-09-651-941-9
4	847	31.7	485	4	US-09-955-597-9
5	847	31.7	508	4	US-09-655-270A-9
6	845	31.6	506	4	US-09-134-001C-4383
7	752	28.1	497	1	US-08-513-841-2
8	752	28.1	497	2	US-08-646-834-2
9	752	28.1	497	2	US-08-942-673-2
10	752	28.1	497	2	US-09-118-317-2
11	731.5	27.4	488	4	US-09-134-001C-4246
12	634	22.7	444	4	US-09-134-001C-4388
13	593.5	22.2	518	4	US-09-134-001C-4451
14	583	21.8	482	4	US-09-155-183-4
15	578.5	21.6	487	4	US-09-351-224E-5
16	380.5	14.2	464	4	US-09-134-001C-4701
17	267	10.0	133	1	US-08-345-611-2
18	267	10.0	133	2	US-08-794-494-2
19	177.5	6.6	551	4	US-08-952-961-2
20	106	4.0	523	4	US-08-613-964-2
21	98.5	3.7	637	1	US-08-235-838-14
22	98.5	3.7	637	2	US-08-465-473B-14
23	97.5	3.6	711	1	US-08-235-838-7
24	97.5	3.6	711	2	US-08-465-473B-7
25	97	3.6	601	1	US-08-333-358-14
26	97	3.6	601	1	US-08-463-694-14
27	97	3.6	601	1	US-08-694-501-14

28	94	3.5	72	4	US-09-655-270A-33	Sequence 33, Appl
29	93.5	3.5	5588	4	US-09-036-987A-6	Sequence 6, Appl
30	93.5	3.5	4584	4	US-09-370-700-6	Sequence 6, Appl
31	92.5	3.5	1257	4	US-08-750-157A-2	Sequence 2, Appl
32	91.5	3.4	484	1	US-08-030-096-8	Sequence 8, Appl
33	90	3.4	793	1	US-09-260-843-2	Sequence 2, Appl
34	90	3.4	343	4	US-09-923-654-2	Sequence 2, Appl
35	89.5	3.3	1485	1	US-07-856-380-2	Sequence 2, Appl
36	89.5	3.3	1385	1	US-07-675-772-2	Sequence 2, Appl
37	89.5	3.3	1385	1	US-08-063-170-2	Sequence 2, Appl
38	89.5	3.3	1385	1	US-08-158-232-2	Sequence 2, Appl
39	89.5	3.3	1385	1	US-08-304-656-2	Sequence 2, Appl
40	89.5	3.3	1385	1	US-08-316-301A-2	Sequence 2, Appl
41	89.5	3.3	1385	2	US-08-611-908-2	Sequence 2, Appl
42	89.5	3.3	1385	2	US-09-173-801-2	Sequence 2, Appl
43	89.5	3.3	1385	4	US-09-076-137-2	Sequence 2, Appl
44	89.5	3.3	1385	5	PCT-5592-03624-2	Sequence 2, Appl
45	89.5	3.3	1385	6	5281530-1	Patent No. 5281530

ALIGNMENTS

RESULT 1

US-09-134-001C-4541

; Sequence 4541, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIORITY APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIORITY FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4541

; LENGTH: 510

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

; US-09-134-001C-4541

Query Match 40.4%; Score 1080; DR 4; Length 510;

Best Local Similarity 43.8%; Prod No 17a-108; Matches 10; Gaps 5;

Matches 215; Conservative 101; Mismatches 167; Pairs 10; Gaps 5;

QY 26 YUNFGGFWAPAGPYQNLTVIGLLGEVASSGKRDIDLADAAHKVKDKWHTSVQ 84

DB 27 YSLTFNNEPLASISSETITVSNPANGPELAKVAVAKKIVTKAVCAAHDAFUSWSKTSKE 86

QY 85 DRAATLFLKLAURMEGNLELLATAETAETWNGKPIRETSADVPDLADHKKYFASCTIACGEG 144

DB 87 EPADYLLLEISREIKETHELATVESLGNKSPYEFTISTIEPFAANQKPYFASVLTDEES 146

QY 145 ISEVSESVYVNHIEPLVWQVITFWNEFLMSKSMAPALAAAGNVCVFAELTFLEVL 204

DB 147 VNEIDUNMSLVNPEVGVGVVAVNNEPILLASWKLPALAAAGNVCVFAELTFLEVL 206

QY 205 LMEIVGDLIPGVVNVVNGAGCVIGFYLATSKRIKAVVATGCTEVQOQIMQVATONIP 264

DB 207 ELAKTEQVLEKGVVNVLTGKSTESNAIPHHGVKLSFSTSTCVYVNAQAIAERIVT 266

QY 265 VTLEGGKSPNIVFADVMDECAFTKALEGFALFAHQCEVTCDFSEALVGFSEYERF 323

DB 267 ITLEGGKSPNIVFADVMDECAFTKALEGFALFAHQCEVTCDFSEALVGFSEYERF 320

QY 324 MEFAIPFVESIPSSNPLSVTQMSALVSHQCEFTTINVTGQKPEFAVITGQPKKLE 382

DB 321 LEKLEAFENIKVSPFPLEDKMSAQTPGPELLKIESYIKIAEEDUKANILTGHP-ITD 379

Db 366 GYYPTPTLFSVTAIMPATRETPVLSVLPFBSEKATILANVWFLAAAVFIPQV 425
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYKSGIGRETHKKMMLHYQQIKCL 500
Db 426 RALRPAQTILACNVWINSWVLPASPYKGGQSGYSGULQAAIESFIKESI 479

RESULT 4

US-09-955-597-9
; Sequence 9, Application US/09955597
; Patent No. 6461856
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E.
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAJNER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09955597
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-955-597-9

Query Match 31.7% Score 847; DB 4; Length 485;
Best Local Similarity 39.0%; Pred. No. 3.5e-83;
Matches 185; Conservative 87; Mismatches 194; Indels 8; Gaps 5;

QY 29 IGCEWVADPGEYCNLTPTVGLLCEVASSGKPDIDLAIAHKYKVKWAHTSVQDPAA 88
Db 12 IGQLTPTSGTATFISINPAQSGHLSVAEATAADVAPAEAAKAAARTWCPMPAPQPTP 71
QY 89 ILFKIADRMFONLELLATAETWNGKPIRETSAADVPALADHFFPYFASCIAPAGEGISEV 148
Db 72 LMFYAAALIEBKTELQLOSPDMGKPIPESLGIDLPIMTLEYFAGLVTKIFGRTPA 131
QY 149 DSETVAYHHEPLGVVGQIIPNPFLLMSKMAPALAAAGNCVVKPARLTPISVLLME 208
Db 132 PGRLNTYLTREPIGVVGAITPNFPAVAVKAPALAMGNAIVIKPAQIAPLVPVALRE 191
QY 209 IVGDL-LPPGVVNVNAGGVIGYVLAISKPIKAVETGSTEVCQIMVATUNILPVIL 267
Db 192 LALEAGLPLGLVNLIPGKSVARNAIVLHPVSVKVFPISTEVQVQIGPMAAURLITASL 251
QY 268 ELGKSPNIVFADVDEEDAFEDKALGFALFAPNQGECVCTCFSRALVQESIYERPMERA 327
Db 252 ELGKKSALVAFGDSPPKAVA---AVVFQAMYS-NUGETCTAPSKLLIVERPIYDEVVELV 306
QY 328 IRVVSISGNPLDSVTOMGAVSHGQLETLINLIDIGKKGALVLLGGPRKLLBELKD 387
Db 307 QARVEAARVGPLODPIETIGPLISAPQRESVHSVSVSDEGSAITLISGQWS-PTSAPEQ 365
QY 388 GYLLPPTILFQON-NMRFQRETEPGVLAVTTEKIMEEALANDTQVIRAGVWSRNGN 446
Db 366 GYYPTPTLFSVTAIMPATRETPVLSVLPFBSEKATILANVWFLAAAVFIPQV 425
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYKSGIGRETHKKMMLHYQQIKCL 500
Db 426 RALRPAQTILACNVWINSWVLPASPYKGGQSGYSGULQAAIESFIKESI 479

RESULT 5

US-09-655-270A-9
; Sequence 9, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouvire, Pierre E.

; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic
; FILE REFERENCE: BC011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/128,792
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-655-270A-9

Query Match 31.7% Score 847; DB 4; Length 508;
Best Local Similarity 39.0%; Pred. No. 3.8e-83;
Matches 185; Conservative 87; Mismatches 194; Indels 8; Gaps 5;

QY 29 IGCEWVADPGEYCNLTPTVGLLCEVASSGKPDIDLAIAHKYKVKWAHTSVQDPAA 88
Db 35 IGQLTPTSGTATFISINPAQSGHLSVAEATAADVAPAEAAKAAARTWCPMPAPQPTP 94
QY 89 ILFKIADRMFONLELLATAETWNGKPIRETSAADVPALADHFFPYFASCIAPAGEGISEV 148
Db 95 LMFYAAALIEBKTELQLOSPDMGKPIPESLGIDLPIMTLEYFAGLVTKIFGRTPA 154
QY 149 DSETVAYHHEPLGVVGQIIPNPFLLMSKMAPALAAAGNCVVKPARLTPISVLLME 208
Db 155 PGRLNTYLTREPIGVVGAITPNFPAVAVKAPALAMGNAIVIKPAQIAPLVPVALRE 214
QY 209 IVGDL-LPPGVVNVNAGGVIGYVLAISKPIKAVETGSTEVCQIMVATUNILPVIL 267
Db 215 LALEAGLPLGLVNLIPGKSVARNAIVLHPVSVKVFPISTEVQVQIGPMAAURLITASL 274
QY 268 ELGKSPNIVFADVDEEDAFEDKALGFALFAPNQGECVCTCFSRALVQESIYERPMERA 327
Db 275 ELGKKSALVAFGDSPPKAVA---AVVFQAMYS-NUGETCTAPSKLLIVERPIYDEVVELV 329
QY 328 IRVVSISGNPLDSVTOMGAVSHGQLETLINLIDIGKKGALVLLGGPRKLLBELKD 387
Db 340 QARVEAARVGPLODPIETIGPLISAPQRESVHSVSVSDEGSAITLISGQWS-PTSAPEQ 388
QY 388 GYLLPPTILFQON-NMRFQRETEPGVLAVTTEKIMEEALANDTQVIRAGVWSRNGN 446
Db 389 GYYPTPTLFSVTAIMPATRETPVLSVLPFBSEKATILANVWFLAAAVFIPQV 448
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYKSGIGRETHKKMMLHYQQIKCL 500
Db 449 RALRPAQTILACNVWINSWVLPASPYKGGQSGYSGULQAAIESFIKESI 502

RESULT 6

US-09-134-001C-4383
; Sequence 4384, Application US/09134001C
; Patent No. 6480370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/364,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/355,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4383
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4383

Query Match 31.6%, Score 845; DB 4; Length 506;
 Best Local Similarity 37.8%; Pred. No. 6; 4-R; 4;
 Matches 190; Conservative 105; Mismatches 184; Indels 24; Gaps 11;

QY 20 KLRKRYDNFVIGGEVWAPADGEYQNLTPVTGQLLCEVASSGKRDIDLDLDAAHKVKD--K 77
 DB 16 KLSNR--QYIDGWEVSSNKKTRDINPNQDETFTVAECTKEDVERAILAARSPEDGE 73
 QY 78 WAHTSVQDRAALFKIADRMQNIJELLATAETWNGKPIRETSAADVPPLAIDHFRYFASC 137
 DB 74 WSLTSEYRGKVRADVAIKENREELAKLETLDGKTLF--SVADMDIHNVMYFAG--131
 QY 138 IRAOEGG-----ISEVDSSETVAVHFPEPLGVVQIILPNWNPILMASWKPAPALACNCV 191
 DB 132 LADKGGELINSPINAEKVV---KEPVGVTQIIPWNYPIIQASWKIAPALATGCSL 187
 QY 192 VIKPARITPELVILLIMETVGL--LPPGVVNVVNGAGGVIGEXLATSKRIAKVAFTGSTEV 250
 DB 188 VMKPEITPLTTIRVFLMERVGPCKGTINVLGASGVGVMSGHEVDLYSFTGGIET 247
 QY 251 GQIIMQVATONIIPVTLFELGKSNIVFADVMDEELAFDKALEGALFAPNGWVCTCP 310
 DB 248 GKHKMQAANVTDVALELGGKPNITTFDA--DFELAVDQALNG---GYFHAGQVCSAG 302
 QY 311 SRALVQESIVYEREMERAIIRVESIRSGNPLDSVTOMCAOVSHGQLETILNYIDICKKEGA 370
 DB 303 SKILVHNDIKKFKKALIDPVSKIKLNGCFQDTFMGPVLISTAHPDKIEGYNEVAKKUGA 362
 QY 371 DVLGTGPRKLEGLKPKDYVILEPTILFG--UNNMVFPVEEIPGHVLAIVTTFKTMEEALELA 429
 DB 363 TTAIGGKRPREF--DLQAGLFFEPVITDCDTSKRIVQEEVFGPVVTFVGFADDEEAIRLA 421
 QY 430 NDTQYGLGAGVMSRNGNLAYKMGRIQAGRWNTNCHAYPAHAAPFGYKQSGIGRTHKM 489
 DB 422 NDSYGLAGALITKIDIGKAORVANKLGLGTWINDHFHPYPAQAPWGKYKQSGIGRELGR 481
 QY 490 MEHYQUTKLLVSYSKPIGLF 512
 DB 482 GLEELYVSKHLLTNTNPEVDWF 504

RESULT 7

US-08-513-841-2

Sequence 2, Application US/08513841

Patent No. 5753481

GENERAL INFORMATION:

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshinori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: No. 5753481el L-sorbose Dehydrogenase and No. 5753481el L-sorbose

TITLE OF INVENTION: Dehydrogenase obtained from Glucosobacter

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/513,841

FILING DATE: 01-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: UK 9304700.9

FILING DATE: 08 MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 241451/1993

FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. ORLON

PRACTICE NUMBER: 24,518

REFERENCE/DOCKET NUMBER: 18-099-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

TELEX: 248855 ODAT OR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 497 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Glucosobacter oxydans

STRAIN: T-100

FEATURE:

NAME/KEY: mat peptide

LOCATION: 1..497

IDENTIFICATION METHOD: experimentally

US-08-513-841-2

Query Match

Best Local Similarity 37.1%; Score 752; DB 1; Length 497;

Matches 187; Conservative 80; Mismatches 197; Indels 40; Gaps 13;

QY 7 SAQIKPEYGFPLKLRKRYDNFVIGGEVWAPADGEYQNLTPVTGQLLCEVASSGKRDIDL 66
 DB 8 SLPLKPREFG-----FIDGWRACKD--FDRSSPAIDVIVTRIPRCTRRLIDE 55
 QY 67 ALDAAHKV--KDKWAHTSVQDRAALFKIADRMQNIJELLATAETWNGKPIRETSAADVP 124
 DB 56 AVAAARAFENGWAGLAAARAAVLLKAAGLLRKRDDIAYWEVLEKCKPSQAG 112
 QY 125 PLADH-----PRYFASCIKAGEG--ISEVDSSETVAVHFPEPLGVVQIILPNWNPILMASW 179
 DB 113 --EIDHICIACFEMAAAGARMJBGDTFNNLGGELFGWLRKPGVGLITTNWDPMTLCE 170
 QY 180 KMADALACNGVVKIPARITPELVILLIMETVGL--LPPGVVNVVNGAGGVIGEXLATSKR 238
 DB 171 RAPPILASGCTIVVKKPAEVTSAITLLARLIALAGLPGKGVENVVIGTGRTVGAMTEHQD 230
 QY 239 IAKVAFTGSTEVQDQ--IMQVATONIIPVTLFELGKSNIVFADVMDEELAFDKALEGALF 297
 DB 231 IDMLSPGTSGVCKSKTHAAADSNIKKHGLELAKKKNPLVVEFALNLEFAA DAVAFG 286
 QY 298 LFAFNQCEVCTCTSRALVQESIVYEREMERAIIRVESIRSGNPLDSVTOMCAOVSHGQLET 357
 DB 287 -LSPNDQGVSSSPALVESVAKELFPLVVTMRPLPGVPLPPLPPLPPLPPLPPLPPLPPL 445
 QY 458 ILNYIDIGKKEGALVLTGGRKELFG ELKQCYLETILEG UNNMVFPVEEIPGHVLAIVTTF 414
 DB 446 ILDYIAKGAEGAKLKGCG -GLVDGKQYLTQPLTFIDVKISMLHARDETFAGVL 399
 QY 415 AVTFTKMEALELANIOTYGLGAGVMSRNGNLAYKMGRIQAGRWNTNCHAYPAHAAPFGY 474
 DB 400 ASHFEDIVDEAIATANDIVYGLAASVMSKIDIKALAVIKRWAGRWNTNCHAYPAHAAP 459
 QY 475 GYKQSGIGRTHKMMLERHYYQTK 498
 DB 460 GGEKQSGWGEAGLYGVEEYTGJK 483

RESULT 8

US-08-696-834-2

Sequence 2, Application US/08696844

Patent No. 5834263

Db 460 GGFKQSGWGRFASIGVVEYTGK 483
 ||:||||| ||| ||: ||| |||

RESULT 11

US-09-134-001C-4246
 ; Sequence 4246, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4246
 ; LENGTH: 488
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4246

Query Match 27.4%, Score 731.5, DR 4, Length 488;
 Best Local Similarity 35.6%, Pred. No. 130-70;
 Matches 175; Conservative 91; Mismatches 204; Indels 21; Gaps 9;

QY	14	EYGPPIKIKARYTNIGRWAPAGGEYQNIIPVTGSLGDEVASSKPOLDALAAHK 74
Db	10	EYN---KMPNPTKQVINGEWVSASSETIDVINGATFVGMKIAKRNPIVKNKAVIAAK 66
QY	74	VKDKWAHTSVQDRAAIIKIALPMQNIELIATAETWNGKPIPETSAATVPLAIDPEY 133
Db	67	VLEPPHSSVEPPRIILKIVKQYGNKNIIEATIDELGAPLSVSNVHYVQMLNHTA 126
QY	134	FASCIQAEGSIGSEVSETVAYHHEPLGVVQIIPNPFILIMASWMAFALAAAGN-VWL 193
Db	127	APALNSQFPFQGRDGLV---KALGVAGVTFNPNFPLNLSKLAAPAFAGSFVL 182
QY	194	KPAITPISVLILMFIIVDEL-LPPQVVVVVWVWVAGSVIYELAYLSKPIAKVAFTOSTEVQ 252
Db	183	KPESETPEFAIIIAEIDFKVGVPGVNLVNGSGVGNPLSEHPKVPKMSFTGSGTGS 242
QY	253	QIMQYALQNIIPVLELNSKSPNIVFALVMPDEE-LAFELKALEGPAIPAPNADEVVTCF 410
Db	243	KIMEKAADFKKVSLELGKSPYIVLDDVDVVEEAAANILKKVN-----NIGQVVIAG 295
QY	311	SRALVQESYIEREMERATIRRVESIRSGNPLDSVTQMGAGVSHGQLETILAYIDIGRKEGA 370
Db	296	TRVLIPESIKEDYLTAVKEAFSKVVGSPREEDVAGSPILSKKQFT-VLDYLUKGINEGA 355
QY	371	DVLIGSKKLLSEELKQGYLEPTILPG-QNNMKVYQEEIEGSPVLAVTTTEKMEAELELA 429
Db	356	ELFYGGPGPK-PEG-LDKGYFARTPIFINVDNHTITAOEEIEGPMVSVITYNNLDEATEIA 413
QY	430	NDTQYGLGAGVSRSPNINLAYKMGGTQAGSPVWTCNYHAYPAHAAGGGYKQSGIGPETHKM 489
Db	414	NTKYLGLAGYVIGKUKOTLKHVARSTEAGTIFINPAGHKP-DLFFGYKSGIGRWGIDY 472
QY	490	MLEHYQCTKCI 500
Db	473	GTEEFLEVKSI 483

RESULT 12

US-09-134-001C-4388
 ; Sequence 4388, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4388
 ; LENGTH: 493
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4388

Query Match 22.6%, Score 684, DR 4, Length 493;
 Best Local Similarity 29.4%, Pred. No. 10-56;
 Matches 145; Conservative 103; Mismatches 204; Indels 42; Gaps 12;

QY	28	FISGEWVAADGEYQNIIPVIGLLCEVASSGKEDIDALDAAHKVKDKWAHTSVQDRA 87
Db	25	FINNEFIESQSKETMDVINPATGEAFDTITLAEENVDALKESSQALKEWKEVPQPTR 84
QY	88	AILPKIADRMEOELLEATAETWNGKPIPETSAADVPILAIIDHFPYFASCIQAEGGI-- 145
Db	85	EHVKLLIPLEKRNDEIAQLVYKEGKTLAQ-AYGEIDKRSISFIDYMTLSMSKGPVLQ 143
QY	146	SEVDETVAHYHHEPLGVVQIIPNPFILIMASWMAFALAAAGN-VWLKARLTPLSVL 205
Db	144	NSIANETIQT-INKVIGVTAQIVPMNAPILVIMKRVIPAVITGVSVMKPESEETILLER 202
QY	206	LMELV-GULLPPEVWVWVWVWVWVAGSVIYELAYLSKPIAKVAFTOSTEVQIMQYALQNIIP 264
Db	203	LAEIPASTIPATLCEIVGKTEVTVTGLASHPTELISLGSMPAGSVYENAAQIVK 262
QY	265	VTELEKSPNIVFALVMPDEE-LAFELKALEGPAIPAPNADEVVTCSPALVORSIYERF 423
Db	253	VNLEGGNAPVIV-----TSNMLEKAVNVIATAPINNAQVQVQPEPIFVHEDVHERF 416
QY	424	MERAFREYESIRSGNPLDSVTQMGAGVSHGQLETILAYIDIGRKEGPAIPAPNADEVVTCSP 383
Db	417	INKVTSKMKSLIVGDFPFPNTDYGALINQKGLDISHHPVQALAKNATLMTGTH----- 370
QY	384	ELK-DEYLEPTILEGNNMP---VFQFFEGEVLAVTTTEKMEFALELANITQYGLGA 438
Db	471	ELKPHQFVYAPVI DNVKPVWVVPKDEIPFVIAITTVYQFQVLELANDINAGLS 427
QY	449	QWSPNPNLAYKMGGTQAGSPVWTC-----YHAYPAHAAGGGYKQSGIGRETHKM 490
Db	428	YIFSENLTVMATPELKEFGEVYANCAEEVWNGYHA-----GWRSGIGGAGDGIHG 479
QY	491	LEHYQCTKLLVSY 504
Db	480	FEYYNTTVSYIRY 493

RESULT 13

US-09-134-001C-4451
 ; Sequence 4451, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4451
 LENGTH: 518
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4451

Query Match 22.2%; Score 593.5; DB 4; Length 518;
 Best Local Similarity 34.0%; Pred. No. 1.5e-55;
 Matches 165; Conservative 8; Mismatches 202; Indels 35; Gaps 16;

QY 20 KIKARYDN---FTGGWVAPADGEYYONLTPV-TGQLICEVASSCKRQIDIALDAARHV 74
 Db 32 KVKAGLQDIPLVINGEKLTQD--TFNSVNPANTSQLIAKYSKATQDDIERAFESANHA 89
 QY 75 KKKWAHTSVQURAAALFKIADRMQNLLELATATWINKPRTSAAADVPLAIDHPRYF 134
 Db 90 YQSKWKSHKDRAEELLIRVAALIRRRKKEISAIMVYFAGKPWDE-AVGDAABGIDIFRY 148
 QY 135 ASCIPAGQGVISFVDSF-TVAYHFHEPLGVVGIIIPWNEPL-LMASWKKMAPALAGNCVV 192
 Db 149 ARSMELADCKPVLDRGEHNRVYKPICTGVTIIPWNEPFAIMAGTTTLAP-VVAGNTVL 207
 QY 193 LKPARLTPSLVLLIMEIVGDL-LPPGVVVVWNGAGGVIGEYLAISKRIAKVAFTGSTVG 251
 Db 208 LKPAEDTVLTAYKIMETILEEAGLPQGVNFVPGDPKREIDGYLDVDRKDTHTFTGSRATG 267
 QY 252 QOI-----MGTATONTIPVTLELGGKSPNIVFADVMDEEDAFFDKALEGFALAFN-QG 304
 Db 268 TRIVERSAVQVQOQFLKRVIAEWGCKDAIVV-----DNNVDTDLAAEAI VTSAGFSG 321
 QY 305 EVCTPSRAIVQESIYEREMERAIRVESIRSNPLDSVTOMGAVSHQGLETLINLID 364
 Db 322 QKCSACSRRAIVHDVHDEILEKAITQKLTGLNTREN-TEMGPVINQKDPKIKNYLEI 380
 QY 365 GKKEGADVLTGGRKLEGLKGYLYEPLILLC-UNNKKVPQELFGVPLAVTTFKIME 423
 Db 381 GKRR-KRTGART-----DSTGYFIETFTSGLOSADRIQOEITFGVVGFIKVKDFD 434
 QY 424 EALELANDTQYIGAGVWSRNCNIAYKMGRTGAGPWFN--CYHAYPAHAAPGKYKQSG 481
 Db 435 EAEIVANDTDTGLTGAIVTHRRCHWKAVNFFDVGNLYINRGCTAAVVGYPHFGGFKMSG 494
 QY 482 IGRET 486
 Db 495 TDART 499

RESULT 14
 US-09-155-183-4
 ; Sequence 4, Application US/09155183
 ; Patent No. 6323011
 ; GENERAL INFORMATION:
 ; APPLICANT: Narbad, Arfan
 ; APPLICANT: Rhodes, Michael J.C.
 ; APPLICANT: Gasson, Michael J.
 ; APPLICANT: Walton, Nicholas J.
 ; TITLE OF INVENTION: PRODUCTION OF VANILLIN
 ; FILE REFERENCE: 20747/100
 ; CURRENT APPLICATION NUMBER: US/09/155, 183
 ; CURRENT FILING DATE: 1999-05-03
 ; EARLIER APPLICATION NUMBER: PCT/GB97/00809
 ; EARLIER FILING DATE: 1997-03-24
 ; EARLIER APPLICATION NUMBER: GB96/06187
 ; EARLIER FILING DATE: 1996-03-23
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas fluorescens
 US-09-155-183-4

Query Match 21.8%; Score 583; DB 4; Length 482;

Best Local Similarity 42.4%; Pred. No. 1.9e-54;
 Matches 158; Conservative 97; Mismatches 189; Indels 44; Gaps 12;

QY 29 IGGFWALADGEYYONLTPVTFGLICEVASSCKRQIDIALDAARHVKKWAHTSVQURAA 88
 Db 8 IGGQSCPARQDRTFERENPVTGELVSKVAATLELDAAVAAAGAAAFPAKAAALNERS 67
 QY 89 ILPKIAURMEONLELATATWINKPRTSAA---DVLALHPRYPASTIRAJE 142
 Db 68 RLKKAAPQDQ-----ARSGEFLAAGTGTAMANNYGFENVRLAANMLKEAASMTLOVN 119
 QY 143 GQISEVQ SETVAHFHFHFGVWQYIPWNEPL-MASWPMALALAAAN-VVLPKAPHTD 201
 Db 120 GFVIPSPVSPSPAMALWQVGVVGLIAPWNAVVIATRALAMPN-ANIVVILKASLSIA 179
 QY 292 SVLLIMEIVGDL-LPPGVVVVWNGAGGVIGEYLAISKRIAKVAFTGSTVGIMY 257
 Db 180 VHRIGQVLDAGLGQGVNVNISNAVDAAQIVERLIANIVRRVNFETSTHGVIVREL 239
 QY 258 ATUNILHVVILLELGRKSHNIVFALVMELEAFLEKALEGLALFA-ENQSEVCTPSKALVO 416
 Db 240 SARHLKALLELGGKRAFLIVL-----DQALFAAVOAAAFGAYNQGDTMSTERLIVD 293
 QY 317 ESIVEREMERAIRVESIRSNPLDSVTOMGAVSHQGLETLINLIDGKKEGALVILG 476
 Db 294 AKVADAFVAGLAAKAVETLRAGVDFESVIGSLIVASA-ETPKALIDUAVAKGKRLVGG 453
 QY 377 KKKLELLELGRKSHNIVFALVMELEAFLEKALEGLALFA-ENQSEVCTPSKALVO 435
 Db 354 Q-----LESTI-----LQPTLLKWDASRLYRERSEFGVAVLGGEGEALQJLANSEFG 405
 QY 436 LGAGVWSRNCNIAYKMGRTGAGVWNTCYHAYT-----AHAFGKYKSGHIGRETIKMM 490
 Db 406 LSAAIFSKULGKALALAAQVSG-----LCHINPVIHDAQWPGGVKSSGSGSGKAS 461
 QY 491 LEHYQOTK 498
 Db 462 IEHFTOLR 469

RESULT 15
 US-09-351-224E-5
 ; Sequence 5, Application US/09351224E
 ; Patent No. 6388171
 ; GENERAL INFORMATION:
 ; APPLICANT: Duvick, Jon
 ; APPLICANT: Maddox, Joyce
 ; APPLICANT: Gilliam, Jacob
 ; APPLICANT: Folkerts, Otto
 ; APPLICANT: Grasta, Oswald R.
 ; TITLE OF INVENTION: Compositions and Methods for Fumonisins
 ; FILE REFERENCE: 5718-111
 ; CURRENT APPLICATION NUMBER: US/09/351, 224E
 ; CURRENT FILING DATE: 1999-07-12
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 487
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-351-224E-5

Query Match 21.6%; Score 578.5; DB 4; Length 487;

Best Local Similarity 33.9%; Pred. No. 5.8e-54;
 Matches 171; Conservative 82; Mismatches 206; Indels 51; Gaps 18;

QY 10 IKPSEYGFELKUKARYINEIGGEHWAPAFAGEYYONLIP-VIQLLEAFVASSAKRQIDIAL 68
 Db 3 LSPDEY-----KSELFINNEFVSSKSGSERLITNTWDESTVATVHVAANAALIVSAV 54
 QY 69 DAAHKV--KDKWAHTSVQURAAALFKIADRMQNLLELATATWINKPRTSAAADVPL 126
 Db 133

Db 55 AASVQAVKKGPWKFTGAQAAACMLKFADLAHNAEKLARLESIPTRPVSMITHFDIPN 114
QY 127 AIDHEFYFASCIRAGGGISEVDSEIVAYHHEPLGVVGOIIPWNFFLLMASWKMAFALA 186
Db 115 MVSFRYYAGWADKIAGKTFFPEDNGKPNWRY-EPMGVCAGIASWNATFLYVGVKIAFALA 173
QY 187 AGNCVVIKPARLTPLSVLLMEIVGDL-LPPCVVNVVNGAGGVIGEVLATSKRIAKVAFT 245
Db 174 AGCSFTFKASEKSPGVIGLAPLFAEAGFPVGVVQFLTGA-PVTGEALASHMDIAKISFT 232
QY 246 GSTEVSQQIMQYAT--QNIIPVILELGGKSNIVFADVMDEDAFFDKALEGFAFALFANQ 303
Db 233 RSVGGGPAVKQ-AILKSNMKFVILELGEK-PTIVNEAPLERQS--GESAKDFSKF---- 284
QY 304 GEVCTCPSPAIWQ-ESIYEPFMEPAIPRVESIPQS-----NPLOSVTQMGQVS 351
Db 285 GQIWWPPSN*LLVQWNIARKF-----HVVHHSFSGN*QFWLQNPFPKPTHGPFVD 436
QY 352 HQOLETILNYIDIGKKEGADVITGGRKLLLEGELKDGYYLEPTILFG-QNNMRVFOEIF 410
Db 337 KSQYDVLGNIDVG-KDTAQALLIGVGRK---GD--KGFALEPTIFVNPCKGSKIWPEEIF 390
QY 411 GPVLAVTTTFTKEEALALELANDTQYGIAGVWSRPNCAIAYKMGEGIQAGRPWTNCYHAYPA 470
Db 391 GPVLSIKTKTEERAEIANDTITGLASVIYTKSLNRLGRVSSALETGGVSINPPFIPET 450
QY 471 HAAFGGYKQSGIGRETHKNMLEHYQQIKCL 500
Db 451 QTPFGQMKQSGSGKELGEBGLKAYLEPKTI 480

Search completed: June 24, 2003, 10:31:44
Job time : 16.1793 secs



Db 67 HAAADAWCKTSVAERALLHRIADRMEEHLEETAVAEETWENGKAVRETTLAADIPALDHE 126

QY 132 RYFASCIQAEQGGISEVDESEIVAYHHEPGLGVGCOIIPWNPPLIMASWKKAPALAGNCV 191

Db 127 RYFAGAIQAEQDSQIDHNTVAYHFNPEPGVGVGCOIIPWNPPLIMATWKLAPALAGNAI 186

QY 192 VLKPARLPLSLVLLMEITVGLDPLPPGVVNVVNGAGGVIGYLYLATSRIKIAKVAFTGSTEVG 251

Db 187 VMKPAEQTPASILYLINIGLDLIDEGVLNIVNGLGRFAGAAALSSNRIRKIAFTGSTEVG 246

QY 252 QOIMQYATONTIIPVTELEGCKSPNIVFADVWDEDAFFDKALEGFALFAFNQGEVCTCPES 311

Db 247 KLINRAASDKIIPVTELEGCKSPSIFESDVLSDQDAFAEKAVEGFAMFALNQGEVCTCPES 306

QY 312 BALVOESYIERFERMERARRVESIRSGNPLDSVTQMGAGVSHGQLEITILNYITFGSKERAP 371

Db 307 RALVHESIADFELEGKVRKVNITKLGNDPLDTETMMGAQASQEQMDKISSYLKIGPEGAG 366

QY 372 VLTGRRKLLCEGLKDRYYIRPTILFGNNMRVQERIFGPVLAIVTFTKMEAEALAN 431

Db 367 TLTGKVKNVQK-MENGYIYPTVPGTNDMPIFPEIFGPVLSVATFSDFDIAIRIAND 425

QY 432 TOYGLGAGVGRNENLAYKMGGRGTOAGRVWNTNCHAYPAHAAPGGYKQSGIGRETHKML 491

Db 426 TNYGLGAGVGRDQNTIYRAGRAQAGRVWNVYHNYPAHSAFGGYKQSGIGRENHMLML 485

QY 492 EHYQOTKCLLYSDKPLGLF 512

Db 486 NHYQOTKCLLYSDNPITGLF 506

RESULT 2

US-09-815-242-10550
 : Sequence 10550, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : FILE REFERENCE: ELITRA.011A
 : CURRENT APPLICATION NUMBER: US/09/815,242
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR APPLICATION NUMBER: 60/191,078
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR APPLICATION NUMBER: 60/206,848
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR APPLICATION NUMBER: 60/207,127
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR APPLICATION NUMBER: 60/242,578
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR APPLICATION NUMBER: 60/253,625
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR APPLICATION NUMBER: 60/257,931
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR APPLICATION NUMBER: 60/269,308
 : PRIOR FILING DATE: 2001-02-16
 : NUMBER OF SEQ ID NOS: 14110
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 10550
 : LENGTH: 496
 : TYPE: prf
 : ORGANISM: Enterococcus faecalis

Query Match

44.4%; Score 1186.5; DB 10; Length 496;

Best Local Similarity 48.6%; Prod. No. 1,90-102;
 Matches 238, Conservative 79; Mismatches 164; Indels 9; Gaps 4;

QY 25 YDNFIFGRHWAPARCFYQYNI TPVTGQI LQFVASSCKRPDIDIAL DAHKVKVKKWAIHVSQ 84

Db 14 YQLTNGFTWTGSGNKKIASYNSGCKLAEPVDAI NADVRAVEAQAQAFQTKRIVDV 74

QY 85 DRAALIFKIDRMFONLELLATATFATUNGKPIRETSAADVPFLADHPRYFASCIKAPES 144

Db 74 TRSNLLIKTADLIEENDEHLMAMVETLNGRPLETUSLIPASADHPRYFASVIRGEES 134

QY 145 ISEVDESEIVAYHHEPGLGVGCOIIPWNPPLIMASWKKAPALAGNCV 204

Db 134 VKFEDKDTLSIVVKEPIGVVGCOIIPWNPPLIMGAKWALATALAAGNIVVHPSSSTSL 194

QY 205 LLMELVGLDPLPPGVVNVVNGAGGVIGYLYLATSRIKIAKVAFTGSTEVGCOIIMQYATONTI 264

Db 194 ELFIPTQVLPKGVVNIITGMSISQSNYMLAHPRFPKIAFTGSTSTFVYTVAKAAADPL 254

QY 265 VTLELGCKSPNIVFADVWDEDAFFDKALEGFALFAFNQGEVCTCPES 424

Db 254 ATLELGCKSANIT EDANKERALEDSVGLILPRAGVVCACGHWVVGGLYDQ 407

QY 324 MERALIRVESIRSGNPLDSVTQMGAGVSHGQLEITILNYITFGSKERKLEES 484

Db 308 VEALKKPEQVNVGFPWEKIVEMCAQINERHOLEELIKYVIRVKEGATILITGQR 466

QY 384 ELKQGYLYLEPTILF-GONNRKVPQEEIFGPVLAIVTFTKMEAEALANITQYGLAGVWS 442

Db 467 GLDKGAPLAPLILANCTNTMCVAQFFIFGPVATVIRKPEFEFVIRLANDSEYGGAGVFS 426

QY 443 RGNLAYKMGGRGTOAGRVWNTNCHAYPAHAAPGGYKQSGIGRETHKMLLEHYQOTKCLIV 502

Db 427 QDINVALKVARGVHGHMWVNYNQLFEASAPFGYKQSGIGRETHKSMIDJAYTOMKNYI 486

QY 503 SYSDKPLGLF 512

Db 487 VTKEEADGLY 496

RESULT 3

US-08-781-986A-5241
 : Sequence 5241, Application US/08781986A
 : Publication No. US20030054436A1
 : GENERAL INFORMATION:
 : APPLICANT: Charles Kunsch
 : TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 : NUMBER OF SEQUENCES: 5255
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Human Genome Sciences, Inc.
 : STREET: 9410 Key West Avenue
 : CITY: Rockville
 : STATE: Maryland
 : COUNTRY: USA
 : ZIP: 20850
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 : COMPUTER: HP Vectra 486/33
 : OPERATING SYSTEM: MSDOS version 6.2
 : SOFTWARE: ASCII Text
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/781,986A
 : FILING DATE:
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Beuson, Bob
 : REGISTRATION NUMBER: 30,446
 : REFERENCE/BOOKET NUMBER: 1B248DP
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (301) 309-8504

```

RESULT 4
US-10-268-518-4
; Subject: 4, Application US/1024R518
; Publication No. US20040106034A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, John Joseph
; TITLE OF INVENTION: 9136. A HYMAN ALDPHYD DEHYDROGNASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: MP101-234PIRM
; CURRENT APPLICATION NUMBER: US/10/268,518
; CURRENT FILING DATE: 2002-10-10
; PCTOP APPLICATION NUMBER: 50/029,899
; PRIOR FILING DATE: 2001-10-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
; US-10-268-518-4

```

RESULT 5

```

: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: 60/211,727
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 21
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: consensus sequence
US-10-175-696-21

```

```

Query Match          35.0%; Score 936; DB 9; Length 493;
Best Local Similarity 45.2%; Pred No 5, 3a-7a;
Matches 225; Conservative 79; Mismatches 160; Indels 34; Gaps 17;

QY 32 EWAPADGEYUNLPVT-GQLCEVASSGKRDIIDLALDA---AHKVKDKWAHTSVODRA 87
DB 1 EWDSASGKTEFVVNPANKGEVIGRVPETAEDVDAVKAAKAEAFKSGPWAKVPASERA 60

QY 88 AILFKIADRMENQLELLATAETWNGKPIRETSA-ALVPLAIUHFYFASCIKAQBG--- 143
DB 61 RILRLADLIEERDELAUETLIDLGKPLAEAGKDTVEGRAIDETIYYAGWARKLMGERR 120

QY 144 ---GISEVDSSETVAYHEHEPLGVVGQIIPWNPFLMASMKMAPALAAAGNCVVLKPARLTP 200
DB 121 VIPSLATDGDDEELNATRRREPLGVVGVIWPNPFLLLALNKLALPALAAGNTVVLKPSQTP 180

QY 201 LSVLLIMELVCDL-----LPGVVVNVNNAKGVIGEVYATSKRIAKVAFSTGVQQLIMJ 256
DB 181 LTALLAELEAGANNLPKGVNVNVPFGAEGVQALLSPDIDKISFTGSTEVGKLIME 240

QY 257 -YATONTIPVTLELGKSPNIVFADVMDEEDAFEDKALEGFALFAF-NOGEVTCPSRAL 314
DB 241 AAAAKNKKVTLGLGKSPVIVF-----DDADLKAVERIVGAFGNAGOVCIAPSRLL 294

QY 315 VOESIYEFMERARIRVESTIR-SCNPLDSVTOM-GAOSVHQULETIL-NYIDIGKREGAD 471
DB 295 VHESIYDFVFKIKERVKKLKGIDPLDSDTNIYGLPISBOQFDRVLSVIFENGKEGAK 354

QY 372 VLTGGRKLLGELGKDYILEPTILFGON-NMVFQFEPFGVPLAVTTTEKTWEAELELAN 430
DB 355 VLCCGERDESKEYLGCGYVYQPTLFTIVTDMKIMKEIFGVPVLPILKFKDLEAELELAN 414

QY 431 DTQYGLGAGVWSRN-GNLAYKMGREGIQAGRWNTN--CYHAYPAHAAPGGYKQ-SGIGRET 486
DB 415 DTEYGLAAYVFTKDIAPAFVAKALFAGIVWVNDVCHAAAPFOLPFGGKQSSGIGRE- 473

QY 487 H-KMMLEHYQOTKCLIV 502
DB 474 HGGKYGLEEYTEIKTVTI 491

```

```

RESULT 6
US-09-823-901-9
: Sequence 9, Application US/09823901
: Patent No. US20020001807A1
: GENERAL INFORMATION:
: APPLICANT: McYers, Rachel
: TITLE OF INVENTION: DEHYDROGENASES AND USES THEREOF
: FILE REFERENCE: 10448-036001
: CURRENT APPLICATION NUMBER: US/09/823,901
: CURRENT FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: 60/194,920
: PRIOR FILING DATE: 2000-03-31
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Artificial Sequence

```

```

: FEATURE:
: OTHER INFORMATION: Consensus Sequence
US-09-823-901-9

Query Match          45.0%; Score 946; DB 10; Length 494;
Best Local Similarity 45.2%; Pred No 5, 3a-7a;
Matches 225; Conservative 79; Mismatches 160; Indels 34; Gaps 17;

QY 32 EWAPADGEYUNLPVT-GQLCEVASSGKRDIIDLALDA---AHKVKDKWAHTSVODRA 87
DB 1 EWDSASGKTEFVVNPANKGEVIGRVPETAEDVDAVKAAKAEAFKSGPWAKVPASERA 60

QY 88 AILFKIADRMENQLELLATAETWNGKPIRETSA-ALVPLAIUHFYFASCIKAQBG--- 143
DB 61 RILRLADLIEERDELAUETLIDLGKPLAEAGKDTVEGRAIDETIYYAGWARKLMGERR 120

QY 144 ---GISEVDSSETVAYHEHEPLGVVGQIIPWNPFLMASMKMAPALAAAGNCVVLKPARLTP 200
DB 121 VIPSLATDGDDEELNATRRREPLGVVGVIWPNPFLLLALNKLALPALAAGNTVVLKPSQTP 180

QY 201 LSVLLIMELVCDL-----LPGVVVNVNNAKGVIGEVYATSKRIAKVAFSTGVQQLIMJ 256
DB 181 LTALLAELEAGANNLPKGVNVNVPFGAEGVQALLSPDIDKISFTGSTEVGKLIME 240

QY 257 -YATONTIPVTLELGKSPNIVFADVMDEEDAFEDKALEGFALFAF-NOGEVTCPSRAL 414
DB 241 AAAAKNKKVTLGLGKSPVIVF-----DDADLKAVERIVGAFGNAGOVCIAPSRLL 294

QY 315 VOESIYEFMERARIRVESTIR-SCNPLDSVTOM-GAOSVHQULETIL-NYIDIGKREGAD 471
DB 295 VHESIYDFVFKIKERVKKLKGIDPLDSDTNIYGLPISBOQFDRVLSVIFENGKEGAK 354

QY 372 VLTGGRKLLGELGKDYILEPTILFGON-NMVFQFEPFGVPLAVTTTEKTWEAELELAN 430
DB 355 VLCCGERDESKEYLGCGYVYQPTLFTIVTDMKIMKEIFGVPVLPILKFKDLEAELELAN 414

QY 431 DTQYGLGAGVWSRN-GNLAYKMGREGIQAGRWNTN--CYHAYPAHAAPGGYKQ-SGIGRET 486
DB 415 DTEYGLAAYVFTKDIAPAFVAKALFAGIVWVNDVCHAAAPFOLPFGGKQSSGIGRE- 473

QY 487 H-KMMLEHYQOTKCLIV 502
DB 474 HGGKYGLEEYTEIKTVTI 491

RESULT 7
US-09-144-882-22
: Sequence 22, Application US/09-144882
: Patent No. US20020162137A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schnable, Patrick S
: APPLICANT: Ke, Jinsban
: APPLICANT: Johnson, Jerry L
: APPLICANT: Allred, Carolyn C
: APPLICANT: Faland, Beth
: APPLICANT: Lutiger, Isabelle
: APPLICANT: Wee, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09/144,882
: CURRENT FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: 03/03,990,717
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: PatentIn Ver. 2.2
: SEQ ID NO 22
: LENGTH: 501
: TYPE: PRT

```


; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

Query Match 34.4%; Score 919; DB 9; Length 501;
Best Local Similarity 39.4%; Pred. No. 2.1e-77;
Matches 196; Conservative 100; Mismatches 165; Indels 36; Gaps 11;

```
QY 28 FIGGEWAPADGEYQNLTPVTQOLCEVASSGRDIDLALDAHAKVKD--KWAHTSVQD 85
DQ 23 FIGQFIDAASCKTFETIDPNGEVIATACGKEDVDLAVNAARAFVCHPWRMTGFE 82
QY 86 RAATLFIKADMEQNLLELATAETWNGKPIRETSAADVPLAIDHFRYFASCIKRAEGGI 145
DQ 83 RAKLTKNFADLLEENIEELAKLDAVDGSKIPLQIGKYADIPATAGHFRYNA-----GAA 135
QY 146 SEVUSETV-----AYHFEHPIGVVQGIIPWNPFLIMASWMAFALAAGNCVVLKPAR 197
DQ 136 DRHGETLKMTPQSLFGYTLKEPIGVVGNITPWNFSPIMEATKVPAPAMAAGCTMVVKPAE 195
QY 198 LPLSVLLMEIVGDL-LPPGVWVNVNAGGVIGEXLATSRIKTAKVAFSTGVGQOIMQ 256
DQ 196 QTSLSALFYAHLKREAGIPDGVNLIVTQGSTAGAAIASHMDVKSFTGSDVGRKIMQ 255
QY 257 -YATONIIPTVLELGGKSPNIVFADVMDEEDAFDDKALEGFALFA--FNOGEVCTCPSPA 313
DQ 256 AAAASNLKKVSLLEGGKSPLLIF-----NDADIDKRAAD-LALLGCFYKNGEICVASSRV 308
QY 314 LVQESIYERFPERAIRRVERISRNGPLDSVTOMGCAQVSHGQLETLINLYIDIGKKGADVL 373
DQ 309 FVQEGIVKVKVEKLVERAKDWTGDPDSTARQGPQVDKQKPKILSYIEHSGNEGATLL 368
QY 374 TGGPKLLEGEFKD-GYYLEPTILFG-QNNMPVQEEIFGVLAVITFKTMBEALAND 431
DQ 369 TGGK-----AIGDKGYIQTIPADVTEDMKIYQDEIFGPVNSLMKFKTVRGKICANN 422
QY 432 TOYGLGAGWBSRNGNLAYKMGREGTOAGRVWNTNCHAYPAHAAFGYKQSGIGRETHKML 491
DQ 423 TRYGLAAGILSQDILINTVSPKIKAGIIWVNYEFGTLLCPYGYKMSGNCPESGMDAL 482
QY 492 EHYQOTKCLLVSYSDKP 508
DQ 483 DNYLQTKSVVPLHNSP 499
```

RESULT 8
US-10-293-865-22

; Sequence 22, Application US/10293865
; Publication No. US20030108000A1

; GENERAL INFORMATION:

; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung

; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants

; FILE REFERENCE: 217113

; CURRENT APPLICATION NUMBER: US/10293865

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/344,882

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 60/090,717

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 22

; LENGTH: 501

; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-22

Query Match 34.4%; Score 919; DB 9; Length 501;
Best Local Similarity 39.4%; Pred. No. 2.1e-77;
Matches 196; Conservative 100; Mismatches 165; Indels 36; Gaps 11;

```
QY 28 FIGGEWAPADGEYQNLTPVTQOLCEVASSGRDIDLALDAHAKVKD--KWAHTSVQD 85
DQ 23 FIGQFIDAASCKTFETIDPNGEVIATACGKEDVDLAVNAARAFVCHPWRMTGFE 82
QY 86 RAAILFKALHMEQNLLELATAETWNGKPIRETSAADVPLAIDHFRYFASCIKRAEGGI 145
DQ 83 RAKLTKNFADLLEENIEELAKLDAVDGSKIPLQIGKYADIPATAGHFRYNA-----GAA 135
QY 146 SEVUSETV-----AYHFEHPIGVVQGIIPWNPFLIMASWMAFALAAGNCVVLKPAR 197
DQ 136 DRHGETLKMTPQSLFGYTLKEPIGVVGNITPWNFSPIMEATKVPAPAMAAGCTMVVKPAE 195
QY 198 LPLSVLLMEIVGDL-LPPGVWVNVNAGGVIGEXLATSRIKTAKVAFSTGVGQOIMQ 256
DQ 196 QTSLSALFYAHLKREAGIPDGVNLIVTQGSTAGAAIASHMDVKSFTGSDVGRKIMQ 255
QY 257 -YATONIIPTVLELGGKSPNIVFADVMDEEDAFDDKALEGFALFA--FNOGEVCTCPSPA 313
DQ 256 AAAASNLKKVSLLEGGKSPLLIF-----NDADIDKRAAD-LALLGCFYKNGEICVASSRV 308
QY 314 LVQESIYERFPERAIRRVERISRNGPLDSVTOMGCAQVSHGQLETLINLYIDIGKKGADVL 373
DQ 309 FVQEGIVKVKVEKLVERAKDWTGDPDSTARQGPQVDKQKPKILSYIEHSGNEGATLL 368
QY 374 TGGPKLLEGEFKD-GYYLEPTILFG-QNNMPVQEEIFGVLAVITFKTMBEALAND 431
DQ 369 TGGK-----AIGDKGYIQTIPADVTEDMKIYQDEIFGPVNSLMKFKTVRGKICANN 422
QY 432 TOYGLGAGWBSRNGNLAYKMGREGTOAGRVWNTNCHAYPAHAAFGYKQSGIGRETHKML 491
DQ 423 TRYGLAAGILSQDILINTVSPKIKAGIIWVNYEFGTLLCPYGYKMSGNCPESGMDAL 482
QY 492 EHYQOTKCLLVSYSDKP 508
DQ 483 DNYLQTKSVVPLHNSP 499
```

RESULT 9

US-09-344-882-20

; Sequence 20, Application US/09344882

; Patent No. US2002012137A1

; GENERAL INFORMATION:

; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung

; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants

; FILE REFERENCE: 201573

; CURRENT APPLICATION NUMBER: US/09/344,882

; CURRENT FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 60/090,717

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.2

; SEQ ID NO 20

; LENGTH: 538

; TYPE: PRT

ORGANISM: Arabidopsis Thaliana
US-09-344-882-20

Query Match 34.3%; Score 918; DB 9; Length 538;
Best Local Similarity 41.1%; Pred. No. 2,96-77;
Matches 202; Conservative 30; Mismatches 170; Indels 30; Gaps 12;

```

QY 29 TCGHWAVADGEYYQNLTPVTGOLLCHVASSGKRDIDIALDAARKVKDK--NAHTSVQUR 86
DB 62 INGNFVSASCKTPTLPPPTGEVIAHVAEEDAEINRAVKAAPTAFADEGPPKMSAYER 121
QY 87 AAILFKIADRMFQNLLELATAETWNGKPIRETSAADVPLAIDHFRVFASCIKRAQEGGTS 146
DB 122 SRVLLRPADLVKHSBELASLETWNGKPYQOSLTAELPMFARLPYACWADKIHGLTI 181
QY 147 EVDSETVAYHHEPLGVVQIIIPWNPFLLMASKMKAPALAAAGNCVVLKPARLTPSLVLL 206
DB 182 PADGNYQVHTLHEPIGVAGQIIIPWNPFLLMFAKVGFPALACGNTIVLKTAEQTLTAFYA 241
QY 207 MEIVGDI-IIPGYNVNVNGAGVIGEYLATSKRTIAKVAFTGSTEVGQOIMQY-ATONITP 264
DB 242 GKLFLEAGLPVGLNIVSGFCATAGALASHMDVDKLAFTGSTDTGKVLGLAANSNLRP 301
QY 265 VTLELGKSPNIVFADWDEDAFDFKALE--GFALFAPNOGEVCTCPSPRALVQESIYER 322
DB 302 VTLELGKSPNIVFADWDEDAFDFKALE--GFALFAPNOGEVCTCPSPRALVQESIYER 322
QY 323 FME-----PAIPPVSTESNPIDSVTQMAGVSNQIIPTLIINYIDGKKEGADVLTGGR 378
DB 355 FVEKSKAPALKRV-----VGDPRKGIPOGQIULKQFVKVYIKSGIESNATLEGG-- 408
QY 379 KLEGELEKQ--GYVLEPIILRG--QNNRVFOEIEFGVLAFTTEKIMEALELANDTOYGL 436
DB 409 ----DQICGKGYFQPTVFSNVKQDMLIAQDEIFGVPQVQSLKFSDDVDVIRANETKYGL 464
QY 437 CAGVWSRNCNLAAYKMGICQAGRWVWNCYHAYPAHAAPGCGYKSGIGRTHKMMLEHYQO 496
DB 465 AAGVETKNLDTANRVSNAIKAGTVVWNCYHAYPAHAAPGCGYKSGIGRTHKMMLEHYQO 496
QY 497 TKCLLVSYSDKP 508
DB 525 IKA-VVIALNKP 535

```

RESULT 10
US-10-293-865-20
Sequence 20, Application US/10293865
Publication No. US2004006090A1
GENERAL INFORMATION:
APPLICANT: Nikolau, Basil J
APPLICANT: Wurtele, Eve S
APPLICANT: Oliver, David J
APPLICANT: Behal, Robert
APPLICANT: Schnable, Patrick S
APPLICANT: Ke, Jushan
APPLICANT: Johnson, Jerry L
APPLICANT: Allred, Carolyn C
APPLICANT: Fatland, Beth
APPLICANT: Lutziger, Isabelle
APPLICANT: Wen, Tsui-Jung
TITLE OF INVENTION: Materials and Methods for the Alteration of: Enzyme and
FILE REFERENCE: 21711
CURRENT APPLICATION NUMBER: US/10/293,865
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/344,882
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 60/040,717
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 20
LENGTH: 538

TYPE: PRT
ORGANISM: Arabidopsis Thaliana
US-10-293-865-20

Query Match 34.4%; Score 918; DB 9; Length 538;
Best Local Similarity 41.1%; Pred. No. 2,96-77;
Matches 202; Conservative 90; Mismatches 170; Indels 30; Gaps 12;

```

QY 29 TCGHWAVADGEYYQNLTPVTGOLLCHVASSGKRDIDIALDAARKVKDK--NAHTSVQUR 86
DB 62 INGNFVSASCKTPTLPPPTGEVIAHVAEEDAEINRAVKAAPTAFADEGPPKMSAYER 121
QY 87 AAILFKIADRMFQNLLELATAETWNGKPIRETSAADVPLAIDHFRVFASCIKRAQEGGTS 146
DB 122 SRVLLRPADLVKHSBELASLETWNGKPYQOSLTAELPMFARLPYACWADKIHGLTI 181
QY 147 EVDSETVAYHHEPLGVVQIIIPWNPFLLMASKMKAPALAAAGNCVVLKPARLTPSLVLL 206
DB 182 PADGNYQVHTLHEPIGVAGQIIIPWNPFLLMFAKVGFPALACGNTIVLKTAEQTLTAFYA 241
QY 207 MEIVGDI-IIPGYNVNVNGAGVIGEYLATSKRTIAKVAFTGSTEVGQOIMQY-ATONITP 264
DB 242 GKLFLEAGLPVGLNIVSGFCATAGALASHMDVDKLAFTGSTDTGKVLGLAANSNLRP 301
QY 265 VTLELGKSPNIVFADWDEDAFDFKALE--GFALFAPNOGEVCTCPSPRALVQESIYER 322
DB 302 VTLELGKSPNIVFADWDEDAFDFKALE--GFALFAPNOGEVCTCPSPRALVQESIYER 322
QY 323 FME-----PAIPPVSTESNPIDSVTQMAGVSNQIIPTLIINYIDGKKEGADVLTGGR 378
DB 355 FVEKSKAPALKRV-----VGDPRKGIPOGQIULKQFVKVYIKSGIESNATLEGG-- 408
QY 379 KLEGELEKQ--GYVLEPIILRG--QNNRVFOEIEFGVLAFTTEKIMEALELANDTOYGL 436
DB 409 ----DQICGKGYFQPTVFSNVKQDMLIAQDEIFGVPQVQSLKFSDDVDVIRANETKYGL 464
QY 437 CAGVWSRNCNLAAYKMGICQAGRWVWNCYHAYPAHAAPGCGYKSGIGRTHKMMLEHYQO 496
DB 465 AAGVETKNLDTANRVSNAIKAGTVVWNCYHAYPAHAAPGCGYKSGIGRTHKMMLEHYQO 496
QY 497 TKCLLVSYSDKP 508
DB 525 IKA-VVIALNKP 535

```

RESULT 11
US-09-961-403-12
Sequence 20, Application US/09961403
Publication No. US20040077589A1
GENERAL INFORMATION:
APPLICANT: HE-STUMPF, HOLGER
APPLICANT: HAEFELDER, BERNARD
APPLICANT: KRAETZSCHMAR, JOERN
APPLICANT: KREFT, BERTHOLD
APPLICANT: WINTERBAGER, ELKE
APPLICANT: REICHER, PEDRO
APPLICANT: SCOTTI, SIMONE
TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
FILE REFERENCE: SCH-1789
CURRENT APPLICATION NUMBER: US/09/961,403
CURRENT FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 512
TYPE: PRT
ORGANISM: Homo sapiens
US-09-961-403-12

Query Match 33.9%; Score 905; DB 9; Length 512;
Best Local Similarity 39.0%; Pred. No. 4,56-76;
Matches 201; Conservative 88; Mismatches 202; Indels 24; Gaps 12;

```

QY 3 NNPPSAQIKPGYGGPLK-LKARYDN-FIGGEWVAPADGEYYQNLTPVTGQLCEVASSG 60
Db 10 NGOPDG--KPPALPRIRMLEVKFTKIFINNHWSKSGKKFATCNSTREQICEVEEGD 67
QY 61 KRIDIALDAAHKVRDK---WAHTSVQDRAAILEKIDMEQNLLELLATATETWNGKPIR 117
Db 68 KPIVIRKAVFAAQVAFQSPWPPIPAISPGPLHQLADLVERDRATLAALLETMDTGKPEL 127
QY 118 ETSAAUPLAIDHFRYFASCIRAOEGGISEVSDSETVYHFEHPGLGVGQIIPWNPFLMA 177
Db 128 HAFIDLEQCFITLRYFAWACKTQCTTPTDDRWVDFTRHEP:GVGCAIFWNPFLML 187
QY 178 SKMAPALAAAGNCVNLKPARCTPLSVLLLMELVGD-LPPGVVNVNVCAGGVIGEYLATS 236
Db 188 VKMLAPALCCGNTWMLKPAEQITPILALYLSIIFKAFPPVAVNVIFGFTVGAATSSH 247
QY 237 KRIKAVFTGSTEVGOQIMQYATQ--NIIPVTLEGGKSPNIVFADVMDEEDAFDDKALEG 295
Db 248 PQINKIAFTGSTEVGKLVKEAASRSNLRVTLLEGGKNPCIVCADA--DIDLAVECAHQG 305
QY 295 FALFAPNGEVTCPSPALVOESIVERPMEPAIRPVESIPSGNPLDSVTOMGAOVSHGOL 355
Db 306 ---VFNGGCCIAASRVEFEVYSEVRKSVYAKKPVGDPPDKTEGPOIDKQF 362
QY 356 ETILNYIDIGKKEGADVLTGGRKLLLEGELKD-GYILEPTILFGQ--NNMRVFOEEIFGP 412
Db 363 DKILEIESGKKEGAKLEGG-----SAMEDKGLFIKPTV-FSEVTDNMRIAKEEIFGP 415
QY 413 VLAVITFKTMEALELANDTQVGLGAGVWSPNGNINLAYPMGRCIAGRWNTNCHYHAYPAHA 472
Db 416 VQILKFKSIEEVIKRANSTDYGLTAAVFTKNLKDALKLASALESGTWINCYNALLYAQA 475
QY 473 AFGSKOSGIGRTHKMLLEHYQOTKCLLVSYSDK 507
Db 476 PFGGFMNSGNRELGEYALAEYTEVKTVTIKLGDK 510

```

RESULT 12

```

US-10-268-518-2
: Sequence 2, Application US/10268518
: Publication NO. US2003010034A1
: GENERAL INFORMATION:
: APPLICANT: Hunter, John Joseph
: TITLE OF INVENTION: 9136, A HUMAN ALGHEHYDE DEHYDROENASE
: TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
: FILE REFERENCE: MP101-234PIRM
: CURRENT APPLICATION NUMBER: US/10/258,518
: CURRENT FILING DATE: 2002-10-10
: PRIOR APPLICATION NUMBER: 60/729,889
: PRIOR FILING DATE: 2001-10-16
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 512
: TYPE: PRT
: ORGANISM: Homo sapien
US-10-268-518-2

```

Query Match 33.9%; Score 905; DB 9; Length 512;

Best Local Similarity 39.0%; Pred. No. 4.5e-76;

Matches 201; Conservative 88; Mismatches 202; Indels 24; Gaps 12;

```

QY 3 NNPPSAQIKPGYGGPLK-LKARYDN-FIGGEWVAPADGEYYQNLTPVTGQLCEVASSG 60
Db 10 NGOPDG--KPPALPRIRMLEVKFTKIFINNHWSKSGKKFATCNSTREQICEVEEGD 67
QY 61 KRIDIALDAAHKVRDK---WAHTSVQDRAAILEKIDMEQNLLELLATATETWNGKPIR 117
Db 68 KPIVIRKAVFAAQVAFQSPWPPIPAISPGPLHQLADLVERDRATLAALLETMDTGKPEL 127
QY 118 ETSAAUPLAIDHFRYFASCIRAOEGGISEVSDSETVYHFEHPGLGVGQIIPWNPFLMA 177
Db 128 HAFIDLEQCFITLRYFAWACKTQCTTPTDDRWVDFTRHEP:GVGCAIFWNPFLML 187

```

```

QY 178 SKMAPALAAAGNCVNLKPARCTPLSVLLLMELVGD-LPPGVVNVNVCAGGVIGEYLATS 236
Db 188 VKMLAPALCCGNTWMLKPAEQITPILALYLSIIFKAFPPVAVNVIFGFTVGAATSSH 247
QY 237 KRIKAVFTGSTEVGOQIMQYATQ--NIIPVTLEGGKSPNIVFADVMDEEDAFDDKALEG 295
Db 248 PQINKIAFTGSTEVGKLVKEAASRSNLRVTLLEGGKNPCIVCADA--DIDLAVECAHQG 305
QY 295 FALFAPNGEVTCPSPALVOESIVERPMEPAIRPVESIPSGNPLDSVTOMGAOVSHGOL 355
Db 306 ---VFNGGCCIAASRVEFEVYSEVRKSVYAKKPVGDPPDKTEGPOIDKQF 362
QY 356 ETILNYIDIGKKEGADVLTGGRKLLLEGELKD-GYILEPTILFGQ--NNMRVFOEEIFGP 412
Db 363 DKILEIESGKKEGAKLEGG-----SAMEDKGLFIKPTV-FSEVTDNMRIAKEEIFGP 415
QY 413 VLAVITFKTMEALELANDTQVGLGAGVWSPNGNINLAYPMGRCIAGRWNTNCHYHAYPAHA 472
Db 416 VQILKFKSIEEVIKRANSTDYGLTAAVFTKNLKDALKLASALESGTWINCYNALLYAQA 475
QY 473 AFGSKOSGIGRTHKMLLEHYQOTKCLLVSYSDK 507
Db 476 PFGGFMNSGNRELGEYALAEYTEVKTVTIKLGDK 510

```

RESULT 13

```

US-10-205-823-14
: Sequence 14, Application US/10205823
: Publication NO. US20030108963A1
: GENERAL INFORMATION:
: APPLICANT: Schlegel, Robert
: APPLICANT: Monahan, John E.
: APPLICANT: Endege, Wilson O.
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Gorbacheva, Bella
: APPLICANT: Hoerscht, Sebastian
: APPLICANT: Kamatkar, Shubhangi
: APPLICANT: Wonsley, Angela M.
: APPLICANT: Glatt, Karen
: APPLICANT: Zhao, Xumel
: APPLICANT: Anderson, Dustin
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
: TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
: FILE REFERENCE: MFI-044
: CURRENT APPLICATION NUMBER: US/10/205,823
: CURRENT FILING DATE: 2002-07-25
: PRIOR APPLICATION NUMBER: 60/407,982
: PRIOR FILING DATE: 2001-07-25
: PRIOR APPLICATION NUMBER: 60/314,356
: PRIOR FILING DATE: 2001-08-22
: PRIOR APPLICATION NUMBER: 60/325,020
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 60/341,746
: PRIOR FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: 60/362,158
: PRIOR FILING DATE: 2002-03-05
: NUMBER OF SEQ ID NOS: 455
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14
: LENGTH: 512
: TYPE: PPT
: ORGANISM: Homo sapiens
US-10-205-823-14

```

Query Match 33.9%; Score 905; DB 9; Length 512;

Best Local Similarity 39.0%; Pred. No. 4.5e-76;

Matches 201; Conservative 88; Mismatches 202; Indels 24; Gaps 12;

```

QY 3 NNPPSAQIKPGYGGPLK-LKARYDN-FIGGEWVAPADGEYYQNLTPVTGQLCEVASSG 60
Db 10 NGOPDG--KPPALPRIRMLEVKFTKIFINNHWSKSGKKFATCNSTREQICEVEEGD 67

```




GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run On: Jun 24, 2003, 10:11:00, Search time: 185 sec, Sequences (without alignments): 1780 862 Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQIKPGVGPPLK ... HYQTKCLIVSNIKIKLGLP 512

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 64473110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database: Pending_Patents_AA_Main.*
- 1: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 7: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 8: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 9: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 10: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 11: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 12: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 13: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
 - 14: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 15: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 16: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 17: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 18: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 19: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 20: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 21: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 22: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 23: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
 - 24: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
 - 25: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
 - 26: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
 - 27: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2668	99.8	512	21	US-09-791-537-8581
2	2668	99.8	512	21	US-09-791-537-9101
3	2668	99.8	512	21	US-09-791-537-9101
4	2668	99.8	512	21	US-09-791-537-9101
5	2668	99.8	512	21	US-09-791-537-9101
6	2668	99.8	512	21	US-09-791-537-9101

7	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
8	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
9	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
10	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
11	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
12	1894	70.9	506	27	US-60-360-039-4614	Sequence 4614, Ap
13	1894	70.9	506	27	US-60-360-039-4614	Sequence 4614, Ap
14	1894	70.9	506	27	US-60-360-039-4614	Sequence 4614, Ap
15	1879	70.3	505	21	US-09-791-537-31772	Sequence 31772, A
16	1879	70.3	505	21	US-09-791-537-31772	Sequence 31772, A
17	1879	70.3	505	21	US-09-791-537-31772	Sequence 31772, A
18	1879	70.3	505	21	US-09-791-537-31772	Sequence 31772, A
19	1868	69.9	493	27	US-60-360-039-4608	Sequence 4608, Ap
20	1868	69.9	493	27	US-60-360-039-4608	Sequence 4608, Ap
21	1868	69.9	493	27	US-60-360-039-4608	Sequence 4608, Ap
22	1856	69.4	495	27	US-60-360-039-17869	Sequence 17869, A
23	1856	69.4	495	27	US-60-360-039-17869	Sequence 17869, A
24	1849.5	69.2	515	21	US-09-791-537-92377	Sequence 92377, A
25	1849.5	69.2	515	21	US-09-791-537-92377	Sequence 92377, A
26	1847	69.1	492	27	US-60-360-039-20615	Sequence 20615, A
27	1823.5	68.2	514	27	US-09-328-352-6997	Sequence 6997, Ap
28	1821	68.1	494	27	US-60-360-039-21741	Sequence 21741, A
29	1810	67.7	506	21	US-09-791-537-32726	Sequence 32726, A
30	1798	67.3	507	21	US-09-791-537-4948	Sequence 4948, Ap
31	1795	67.2	682	16	US-09-252-992A-31897	Sequence 31897, A
32	1789	66.9	506	21	US-09-791-537-10758	Sequence 10758, A
33	1783	66.7	506	27	US-60-360-039-8796	Sequence 8796, Ap
34	1780.5	66.6	505	21	US-09-791-537-11175	Sequence 11175, A
35	1773	66.3	506	27	US-60-360-039-14000	Sequence 14000, A
36	1768.5	66.2	506	21	US-09-791-537-23318	Sequence 23318, A
37	1760	65.8	507	21	US-09-791-537-68900	Sequence 68900, A
38	1759	65.8	534	18	US-09-489-039A-10148	Sequence 10148, A
39	1714.5	64.1	506	20	US-09-602-740-218	Sequence 218, App
40	1714.5	64.1	506	20	US-09-602-740-218	Sequence 218, App
41	1714.5	64.1	506	20	US-09-602-740-218	Sequence 218, App
42	1682	62.9	506	21	US-09-791-537-6572	Sequence 6572, Ap
43	1682	62.9	506	21	US-09-791-537-6572	Sequence 6572, Ap
44	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
45	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
46	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
47	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
48	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
49	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl
50	1595.5	58.6	405	14	US-09-459-856-23	Sequence 23, Appl

ALIGNMENTS

US-09-791-537-8581
Sequence 8581, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Biocomix, Inc.
APPLICANT: Debe, Derek
APPLICANT: Danzer, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 241/210
CURRENT APPLICATION NUMBER: US/09/791-537
NUMBER OF SEQ ID NOS: 154055
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8581
LENGTH: 512
TYPE: PRT
ORGANISM: Escherichia coli

US-09-791-537-8581

Query Match 99.8%, Score 2668, DB 21, Length 512,
Best Local Similarity 94.8%, Prd No 1, Pos 265,
Matches 511, Conservative 0, Mismatches 1, Indels 0, Gaps 0;

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||
1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

1 MTNPPSAQIKPGVGPPLKARYDNFPGWVAPADBYVYONLTPVRSQGLCEVASSG 60
|||||

Db 61 KPDIDALDAHRAHKVKKWAHTSVQDRAAIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 120
QY 121 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 180
Db 121 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 180
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 240
QY 241 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 300
Db 241 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 300
QY 301 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 360
Db 301 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 360
QY 361 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 420
Db 361 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 420
QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGKGIQAGRVWTCYHAYPAHAAPFGYKQS 480
Db 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGKGIQAGRVWTCYHAYPAHAAPFGYKQS 480
QY 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512
Db 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512

RESULT 2

US-09-791-537-9101

: Sequence 9101, Application US/09791537

: GENERAL INFORMATION:

: APPLICANT: Biomimix, Inc.

: APPLICANT: Debe, Derek

: APPLICANT: Danzer, Joseph

: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

: TITLE OF INVENTION: METHODS OF USE THEREOF

: FILE REFERENCE: 261/210

: CURRENT APPLICATION NUMBER: US/09/791,537

: CURRENT FILING DATE: 2001-02-22

: NUMBER OF SEQ ID NOS: 153055

: SOFTWARE: Patent In version 3.0

: SEQ ID NO 9101

: LENGTH: 542

: TYPE: PRT

: ORGANISM: Escherichia coli

US-09-791-537-9101

Query Match 99.8%; Score 2668; DB 21; Length 542,
Best Local Similarity 99.8%; Pred. No. 1.6e 265;
Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPEYGFPLKARYDNFTGGVWVAPADGEYYONITPTVTGQLCEVASSG 60
Db 31 MTNNPSSAQIKPEYGFPLKARYDNFTGGVWVAPADGEYYONITPTVTGQLCEVASSG 90
QY 61 KRDIIDLDAAHKVKDKWAHTSVQDRAAIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 120
Db 91 KRDIIDLDAAHKVKDKWAHTSVQDRAAIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 150
QY 121 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 180
Db 151 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 210
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 270
QY 241 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 300

Db 271 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 440
QY 301 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 460
Db 331 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 490
QY 361 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 420
Db 391 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 450
QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGKGIQAGRVWTCYHAYPAHAAPFGYKQS 480
Db 451 TMEALELANDTOYGLGAGVWSRNGNLAYKMGKGIQAGRVWTCYHAYPAHAAPFGYKQS 510
QY 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512
Db 511 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 542

RESULT 3

US-09-340-049-23692

: Sequence 23692, Application US/00360049

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei

: APPLICANT: Chen, Xianfeng

: APPLICANT: Goldman, Barry S.

: APPLICANT: Hinkle, Gregory J.

: APPLICANT: Slater, Steven C.

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 48-10(52052)A

: CURRENT APPLICATION NUMBER: US/60/360,049

: CURRENT FILING DATE: 2002-02-21

: NUMBER OF SEQ ID NOS: 47374

: SEQ ID NO 23692

: LENGTH: 542

: TYPE: PRT

: ORGANISM: Escherichia coli

US-09-360-039-23692

Query Match 99.8%; Score 2648; DB 27; Length 542,
Best Local Similarity 99.8%; Pred. No. 1.6e 265;
Matches 511; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPEYGFPLKARYDNFTGGVWVAPADGEYYONITPTVTGQLCEVASSG 60
Db 31 MTNNPSSAQIKPEYGFPLKARYDNFTGGVWVAPADGEYYONITPTVTGQLCEVASSG 90
QY 61 KRDIIDLDAAHKVKDKWAHTSVQDRAAIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 120
Db 91 KRDIIDLDAAHKVKDKWAHTSVQDRAAIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 150
QY 121 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 180
Db 151 AADVPLAIDHFRFYFASCIRAEQGGISEVDSEIVAYHFHEPLGVVGGQIIPWNPFLLMASWK 210
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGGVIGEYATSKRIA 270
QY 241 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 400
Db 271 KVAFTGSTVGQOIMOVATONIIIPVTLLELGGKSPNIVFADVMDEDAEDFEDKALEGFALFA 440
QY 301 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 460
Db 331 FNOGEVCTPSRALVQESIVYERFMERAIIRKVESIRSGNPLDSVTOMGADVSHQLETLIN 490
QY 361 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 420
Db 391 YIDIGKKEGADVLTCGRRKILLEGELKDGYYLEPTILFGONNMNVFQEEIFGVLAVTTFK 450

QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 480
 DB 451 TMFAAIFANIQYGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 510
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 DB 511 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 542

RESULT 4

US-09-252-691-9975
 : Sequence 9975, Application US/09252691B
 : GENERAL INFORMATION:
 : APPLICANT: Keith G. Weinstein et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
 : FILE REFERENCE: 107196.135
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US/09/252.691B
 : NUMBER OF SEQ ID NOS: 11324
 : SEQ ID NO 9975
 : LENGTH: 546
 : TYPE: PRT
 : ORGANISM: Enterobacter cloacae
 US-09-252-691-9975

Query Match 95.88; Score 2560; DB 16; Length 546;
 Best Local Similarity: 94.3%; Pred. No. 2.4e-254;
 Matches 483; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNNPPSAGIKPGEGYFPIKIKAPYDNFIFGFWVAPADGYQNIPTVTGQILCEVASS 60
 DB 35 MTNNPPSRIQGEYGFPIKIPRYNFIRSLWVAPVSGYSNLIPVTGQILCEVASS 94
 QY 61 KRDIIDALDAAHKAKKQKQTSVQRAALIFKIDRPEQNLLELLATAETWONKPIRET 120
 DB 95 KRDIIDALDAAHKAKKQKQTSVQRAALIFKIDRPEQNLLELLATAETWONKPIRET 154
 QY 121 AADVPLAIDHFRYFASCIKRAQEGGISEVDSETVAYHHEPLGVVGGQIIPWNPFLMASWK 180
 DB 155 AADVPLAIDHFRYFASCIKRAQEGGISEVDKDTVAYHHEPLGVVGGQIIPWNPFLMASWK 214
 QY 181 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDILLPGCVINVVNGAGGEGEYLATSKRIA 240
 DB 215 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDILLPGCVINVVNGAGGEGEYLATSKRIA 274
 QY 241 KVAFTGSTEVGQIQMAYATONIIPTVLELGKSPNIVFADVMDDEDAFFDKALEGFALFA 300
 DB 275 KVAFTGSTEVGQIQMAYATONIIPTVLELGKSPNIFADYMDDEDAFFDKALEGFALFA 334
 QY 301 FNQGEVCTCPSRALVQESYIERKPMERAIKRVESIPSNIPISVTQMGAAVSHQOLETILN 360
 DB 335 FNQGEVCTCPSRALVQESYIERKPMERAIKRVESIPSNIPISVTQMGAAVSHQOLETILN 394
 QY 361 YIDIGKKRGAIVLGGRRKKVLEGEIKQGYLPTLPGQNNMRVFOEIPGPVLAVTTTK 420
 DB 395 YIDIGKKRGAIVLGGRRKKVLEGEIKQGYLPTLPGQNNMRVFOEIPGPVLAVTTTK 454
 QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 480
 DB 455 TMFAAIFANIQYGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 514
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 DB 511 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 546

RESULT 5

US-09-252-691C-9975
 : Sequence 9975, Application US/09252691C
 : GENERAL INFORMATION:
 : APPLICANT: Keith G. Weinstein et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

: TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196.135
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US 60/094,145
 : PRIOR FILING DATE: 1998-07-24
 : PRIOR APPLICATION NUMBER: US 60/074,787
 : PRIOR FILING DATE: 1998-02-18
 : NUMBER OF SEQ ID NOS: 11326
 : SEQ ID NO 9975
 : LENGTH: 546
 : TYPE: PRT
 : ORGANISM: Enterobacter cloacae
 US-09-252-691C-9975

Query Match 95.88; Score 2560; DB 16; Length 546;
 Best Local Similarity: 94.3%; Pred. No. 2.4e-254;
 Matches 483; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNNPPSAGIKPGEGYFPIKIKAPYDNFIFGFWVAPADGYQNIPTVTGQILCEVASS 60
 DB 35 MTNNPPSRIQGEYGFPIKIPRYNFIRSLWVAPVSGYSNLIPVTGQILCEVASS 94
 QY 61 KRDIIDALDAAHKAKKQKQTSVQRAALIFKIDRPEQNLLELLATAETWONKPIRET 120
 DB 95 KRDIIDALDAAHKAKKQKQTSVQRAALIFKIDRPEQNLLELLATAETWONKPIRET 154
 QY 121 AADVPLAIDHFRYFASCIKRAQEGGISEVDSETVAYHHEPLGVVGGQIIPWNPFLMASWK 180
 DB 155 AADVPLAIDHFRYFASCIKRAQEGGISEVDKDTVAYHHEPLGVVGGQIIPWNPFLMASWK 214
 QY 181 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDILLPGCVINVVNGAGGEGEYLATSKRIA 240
 DB 215 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDILLPGCVINVVNGAGGEGEYLATSKRIA 274
 QY 241 KVAFTGSTEVGQIQMAYATONIIPTVLELGKSPNIVFADVMDDEDAFFDKALEGFALFA 300
 DB 275 KVAFTGSTEVGQIQMAYATONIIPTVLELGKSPNIFADYMDDEDAFFDKALEGFALFA 334
 QY 301 FNQGEVCTCPSRALVQESYIERKPMERAIKRVESIPSNIPISVTQMGAAVSHQOLETILN 360
 DB 335 FNQGEVCTCPSRALVQESYIERKPMERAIKRVESIPSNIPISVTQMGAAVSHQOLETILN 394
 QY 361 YIDIGKKRGAIVLGGRRKKVLEGEIKQGYLPTLPGQNNMRVFOEIPGPVLAVTTTK 420
 DB 395 YIDIGKKRGAIVLGGRRKKVLEGEIKQGYLPTLPGQNNMRVFOEIPGPVLAVTTTK 454
 QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 480
 DB 455 TMFAAIFANIQYGLGAGVWSRNLAYKMGRIQAGRYWTCYHAYPAHAFAFGYKOS 514
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 DB 511 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 546

RESULT 6

US-60-360-039-17072
 : Sequence 17072, Application US/60360039
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yongwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : FILE REFERENCE: 38-10,52052A
 : CURRENT APPLICATION NUMBER: US/60/360,039
 : CURRENT FILING DATE: 2002-02-21
 : NUMBER OF SEQ ID NOS: 47374
 : SEQ ID NO 17072
 : LENGTH: 506

```

; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-60-360-039-17072

Query Match      71.8%; Score 1918; DB 27; Length 506;
Best Local Similarity 71.7%; Pct-Id No. 4, 4e-188;
Matches 352; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 22 KARYDNFIGGWWAPADGEYQNLTPVTGQLCEVASSGKRDIDLALDAHKKVKDKWAHT 81
DB 16 KARYDNEIGGWWAPADGRYFNSSPLHGKIKCEIARSQAIDIERALDAHAAKAGWART 75
QY 82 SVQDRAAIIKFIADRMFONIELLATAETWNGKPIRETSAADVPLAIDHRYFASCTRAQ 141
DB 76 SAADRSNIIIPRIAMFEENIAALATAETWNGKPIRETAAIPLAIDHRYFASCTRAQ 135
QY 142 EGGISEVDSETVAYHHEPGLGVVGGQIIPWNPFLLMASWMAFALAAANVVLKPAETPL 201
DB 136 EGGISEIDHDTIAYHHEPGLGVVGGQIIPWNPFLLMACWKALPALAAGNCVVLKPAEQTPA 195
QY 202 SVLLMEIVGDLPPGVVNVVNGAGGVGEYLAISKRIAKVAFTSTEVGGQIMOVATON 261
DB 196 SIMWAMIGDLPAGVNLVINGFLEAGKPLASSPRIAKIAPTGETSTGRLLIMOYAAQN 255
QY 262 IIPVTLGLGKSPNIVFADYMDDEDAFFDKALGFALFAPNOCVECTCPSRALVQESIVE 321
DB 256 IIPVTLGLGKSPNIFEDDVAREDDYLDKALEGFTMFALNOGEVCTCPSRALVQESIVE 315
QY 322 RFMERARRVSTRSNPLDSVTOMGAQVSHGOLEITILAYIDIGKKEGADVITGGRRKLL 381
DB 316 KFMERALKRVNAVQGSPLDPAIMIGAQAASEQNLKILGYMDIGRNEGAKLLAGQRRKIL 375
QY 382 EGELKDGYYLEPTILFQNNMRVFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 441
DB 376 PGQLADGYVEPTEVFEHGNKMRIFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 435
QY 442 SRNCLAYKMGRIQAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 501
DB 436 SRDANCYRFRGRIEAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 495
QY 502 VSYSDKPLGLF 512
DB 496 VSYSPKALGFF 506

RESULT 7
US-60-360-039-15459
; Sequence 15459, Application US/60-360-039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianteng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15459
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-60-360-039-15459

Query Match      71.7%; Score 1917; DB 27; Length 495;
Best Local Similarity 71.9%; Pct-Id No. 5, 4e-188;
Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

QY 22 KARYDNFIGGWWAPADGEYQNLTPVTGQLCEVASSGKRDIDLALDAHKKVKDKWAHT 81
DB 5 KPRYDNFIGGWWAPADGQGYFNTPTTGKVTSTARSTAPPIDIAALDAHAAKFAWCKT 64
QY 82 SVQDRAAIIKFIADRMFONIELLATAETWNGKPIRETSAALVPLAIDHRYFASCTRAQ 141
DB 65 STTDRSNVLLKIADRIEQNLLELAYAFTWNGKPKVRETINAIIVPLCVIHFHYFAAATPAQ 124
QY 142 EGGISEVDSETVAYHHEPGLGVVGGQIIPWNPFLLMASWMAFALAAANVVLKPAETPL 201
DB 126 EGGISEIDSDTIAYHHEPGLGVVGGQIIPWNPFLLMASWMAFALAAANVVLKPAETPL 184
QY 202 SVLLMEIVGDLPPGVVNVVNGAGGVGEYLAISKRIAKVAFTSTEVGGQIMOVATON 261
DB 196 SIMWAMIGDLPAGVNLVINGFLEAGKPLASSPRIAKIAPTGETSTGRLLIMOYAAQN 255
QY 262 IIPVTLGLGKSPNIVFADYMDDEDAFFDKALGFALFAPNOCVECTCPSRALVQESIVE 321
DB 256 IIPVTLGLGKSPNIFEDDVAREDDYLDKALEGFTMFALNOGEVCTCPSRALVQESIVE 315
QY 322 RFMERARRVSTRSNPLDSVTOMGAQVSHGOLEITILAYIDIGKKEGADVITGGRRKLL 381
DB 316 KFMERALKRVNAVQGSPLDPAIMIGAQAASEQNLKILGYMDIGRNEGAKLLAGQRRKIL 375
QY 382 EGELKDGYYLEPTILFQNNMRVFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 441
DB 376 PGQLADGYVEPTEVFEHGNKMRIFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 435
QY 442 SRNCLAYKMGRIQAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 501
DB 436 SRDANCYRFRGRIEAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 495
QY 502 VSYSDKPLGLF 512
DB 496 VSYSPKALGFF 506

RESULT 8
US-60-360-039-15827
; Sequence 15827, Application US/60-360-039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianteng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15827
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-60-360-039-15827

Query Match      71.7%; Score 1917; DB 27; Length 495;
Best Local Similarity 71.9%; Pct-Id No. 5, 4e-188;
Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

QY 22 KARYDNFIGGWWAPADGEYQNLTPVTGQLCEVASSGKRDIDLALDAHKKVKDKWAHT 81
DB 5 KPRYDNFIGGWWAPADGQGYFNTPTTGKVTSTARSTAPPIDIAALDAHAAKFAWCKT 64
QY 82 SVQDRAAIIKFIADRMFONIELLATAETWNGKPIRETSAALVPLAIDHRYFASCTRAQ 141
DB 65 STTDRSNVLLKIADRIEQNLLELAYAFTWNGKPKVRETINAIIVPLCVIHFHYFAAATPAQ 124
QY 142 EGGISEVDSETVAYHHEPGLGVVGGQIIPWNPFLLMASWMAFALAAANVVLKPAETPL 201
DB 126 EGGISEIDSDTIAYHHEPGLGVVGGQIIPWNPFLLMASWMAFALAAANVVLKPAETPL 184
QY 202 SVLLMEIVGDLPPGVVNVVNGAGGVGEYLAISKRIAKVAFTSTEVGGQIMOVATON 261
DB 196 SIMWAMIGDLPAGVNLVINGFLEAGKPLASSPRIAKIAPTGETSTGRLLIMOYAAQN 255
QY 262 IIPVTLGLGKSPNIVFADYMDDEDAFFDKALGFALFAPNOCVECTCPSRALVQESIVE 321
DB 256 IIPVTLGLGKSPNIFEDDVAREDDYLDKALEGFTMFALNOGEVCTCPSRALVQESIVE 315
QY 322 RFMERARRVSTRSNPLDSVTOMGAQVSHGOLEITILAYIDIGKKEGADVITGGRRKLL 381
DB 316 KFMERALKRVNAVQGSPLDPAIMIGAQAASEQNLKILGYMDIGRNEGAKLLAGQRRKIL 375
QY 382 EGELKDGYYLEPTILFQNNMRVFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 441
DB 376 PGQLADGYVEPTEVFEHGNKMRIFOEIEFGPVLAVTTKTMEAEALFLANDTOYGLGAGVW 435
QY 442 SRNCLAYKMGRIQAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 501
DB 436 SRDANCYRFRGRIEAGRVWNTCYHAYPAHAAPGGYKQSGIGKRETHKMMLEHYOOTKCL 495
QY 502 VSYSDKPLGLF 512
DB 496 VSYSPKALGFF 506

```

```

Db 185 SILVMEVIGDLPPGVNVYNGFGLGKPKLASNPRIAKTFTGTTTGTIMOVASON 244
QY 262 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 321
Db 245 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 304
QY 322 REMEAIKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 364
Db 305 TWEKALKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 364
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGPGVLAVTTFTKMEAELELANDTQYGLGAGVW 441
Db 365 DGEISDGFYVKTPEFKGHKMFVQEEIFGPGVWVTTFKDEADALAIANDTLYGLGAGVW 424
QY 442 SNGNLAYKMGAGIUGAVWNTNCHAYPAHAAPFSGYKUSGIGRETHKMWLFPHVQOTKCLL 501
Db 425 SDASRLYRMGRALQAGRWNTNCHAYPAHAAPFSGYKUSGIGRETHKMWLFPHVQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 9
US-60-360-039-16210
; Sequence 16210, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16210
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-60-360-039-16210

```

```

Query Match 71.7%; Score 1917; DB 27; Length 495;
Best Local Similarity 71.9%; Pred No. 5.4e-188;
Matches 353; Conservative 58, Mismatches 80, Indels 0; Gaps 0;

QY 22 KARYDNFIIGGEWAFADGEYYQNLTFVTGQLLCEVASSGKRDIDLDAAHKVKDKWAHT 81
Db 5 KRYDNFIIGGEWAFADGEYYQNLTFVTGQLLCEVASSGKRDIDLDAAHKVKDKWAHT 64
QY 82 SVQDRAALFKIADRMENLELATAETWNGKPIRETSAADVPFLAIDHFERYFASCIQAQ 141
Db 65 STIDRSNVLLKTIADRIEALNLELATAETWNGKPIRETSAADVPFLAIDHFERYFASCIQAQ 124
QY 142 EGGISEVDSETVAYHFHEPLGVVGQIIPWNEPFLMASWKMALPAAACNVVLPKAPLTP 201
Db 125 EGGISEVDSETVAYHFHEPLGVVGQIIPWNEPFLMASWKMALPAAACNVVLPKAPLTP 184
QY 202 SVLLMEIVGDLPPGVNVYNGFGLGKPKLASNPRIAKTFTGTTTGTIMOVASON 261
Db 185 SILVMEVIGDLPPGVNVYNGFGLGKPKLASNPRIAKTFTGTTTGTIMOVASON 244
QY 262 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 321
Db 245 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 304
QY 322 REMEAIKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 381
Db 305 TWEKALKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 364
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGPGVLAVTTFTKMEAELELANDTQYGLGAGVW 441

```

```

Db 365 DGEISDGFYVKTPEFKGHKMFVQEEIFGPGVWVTTFKDEADALAIANDTLYGLGAGVW 424
QY 442 SNGNLAYKMGAGIUGAVWNTNCHAYPAHAAPFSGYKUSGIGRETHKMWLFPHVQOTKCLL 501
Db 425 SDASRLYRMGRALQAGRWNTNCHAYPAHAAPFSGYKUSGIGRETHKMWLFPHVQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 10
US-09-791-537-8001
; Sequence 8001, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Debe, Derek
; APPLICANT: Dancer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8001
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Ralstonia eutropha
US-09-791-537-8001

```

```

Query Match 71.5%; Score 1910; DB 21; Length 506;
Best Local Similarity 71.5%; Pred No. 4e-187;
Matches 351; Conservative 67, Mismatches 73, Indels 0; Gaps 0;

QY 22 KARYDNFIIGGEWAFADGEYYQNLTFVTGQLLCEVASSGKRDIDLDAAHKVKDKWAHT 81
Db 16 KQYENYIGGAWVPPAGGEYFESTPTITGKPTFVRSGOODVDAALDAHAAKAAWART 75
QY 82 SVQDRAALFKIADRMENLELATAETWNGKPIRETSAADVPFLAIDHFERYFASCIQAQ 141
Db 76 STTERANILNRIADRIEALNLELATAETWNGKPIRETSAADVPFLAIDHFERYFASCIQAQ 135
QY 142 EGGISEVDSETVAYHFHEPLGVVGQIIPWNEPFLMASWKMALPAAACNVVLPKAPLTP 201
Db 136 EGGISEVDSETVAYHFHEPLGVVGQIIPWNEPFLMASWKMALPAAACNVVLPKAPLTP 195
QY 202 SVLLMEIVGDLPPGVNVYNGFGLGKPKLASNPRIAKTFTGTTTGTIMOVASON 261
Db 196 SILVMEVIGDLPPGVNVYNGFGLGKPKLASNPRIAKTFTGTTTGTIMOVASON 255
QY 262 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 321
Db 256 IIPVILEGGKSPNIVFAUWDEEDAFEDKALGEGFALFAFNOGEVCHTCSBALVQESIYE 315
QY 322 REMEAIKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 381
Db 316 REMEAIKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 375
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGPGVLAVTTFTKMEAELELANDTQYGLGAGVW 441
Db 376 DGEISDGFYVKTPEFKGHKMFVQEEIFGPGVWVTTFKDEADALAIANDTLYGLGAGVW 435
QY 442 SNGNLAYKMGAGIUGAVWNTNCHAYPAHAAPFSGYKUSGIGRETHKMWLFPHVQOTKCLL 501
Db 446 TWEKALKRVKRAIKOGNPLDNTVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 495
QY 502 VSYSDKPLGLF 512
Db 496 VSYSPKALGFF 506

```

RESULT 11

US-60-360-039-12218
 : Sequence 12218, Application US/60360039
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yonqwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : FILE REFERENCE: 38-10(52052)A
 : CURRENT APPLICATION NUMBER: US/60/360,039
 : CURRENT FILING DATE: 2002-02-21
 : NUMBER OF SEQ. ID NOS: 47374

SEQ. ID NO 12218

LENGTH: 505

TYPE: PRT

ORGANISM: Mesorhizobium loti

US-60-360-039-12218

Query Match: 71.3%; Score 1905; DB 27; Length 505;

Best Local Similarity 70.3%; Pred. No. 9,66-187;

Matches 352; Conservative 63; Mismatches 84; Indels 2; Gaps 1;

QY 14 EYGFPLK--LKARYDNF IGGEWVAPADGEYVQNLTPVTGQLLCEVASSGKRKIDLDLDA 71

Db 5 EFSRPVAPKPKRYGNF IGKKWTEPRSGRYFNHSPVQQLCEVARSDDADDEAALDA 64

QY 72 HKYDKWAHTSVQDRAALFKIADRMFQNLLELTAETWNGKPIRETSAALVPIADHF 131

Db 65 HAAKDAGRTSVASRAILNRIADRMFENLDAETWNGKPIRETIVADPLIADHF 124

QY 132 RYFASCIQAEGGISEVDESETVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 191

Db 125 RYFASVYGGQPSQLSDDDTVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 184

QY 192 VLKPARUTPLSVLLMEIVGILLPPGVVNVVNGAGVIGEYIATSKRIAKVAFSTGVG 251

Db 185 VLKPAEUTPAILLWADLIGDLLPGVLNVNGFLEAGKPLASSPRIAKIATGETTIG 244

QY 252 QQIMGVATONIIPVTELEGGKSPNIVFAVMDEEDAFDALKLEGFALFAFNGVCVCS 311

Db 245 RLIMVASQNLIPVTELEGGKSPNIFKDVVAEDGGDFDALKLEGVMEALNAGEVCVCS 304

QY 312 RALVQESIYEFMEPAIRPVESTIRSGNPLDSVTOMGAQVSHGOLETIILNYIDIGKKEGAD 371

Db 305 RALIHESIYDFMERALKRVAEIVQGDPLDPATMICAQASSPQLEKILSYIDIGHQEGAE 364

QY 372 VLTGGRKLEGELEKDCYYLEPTILFGQNNMRVFOEEIFGPVLAVTTFTKMEALELAND 431

Db 365 VLTGGAFNVLPGLAGSYVYKPTVFRGNHNMRIFOEEIFGPVVSVTTFKDDDEALSAND 424

QY 432 TOYGLGAVMSRGNLAYKMGRIQAGRWNTNCHYATPAHAAFGYKQSGIGKRTHKMML 491

Db 425 TLYGLGAVWTRCNRAYRFGRAIQAGRWNTNCHYATPAHAAFGYKQSGIGRETHKML 484

QY 492 EHYQOTKCLLVSYSDKPLGLF 512

Db 485 DRYQOTKMLVSYSPKKLGLFF 505

RESULT 12

US-60-360-039-4614
 : Sequence 4614, Application US/60360039
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yonqwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : FILE REFERENCE: 38-10(52052)A
 : CURRENT APPLICATION NUMBER: US/60/360,039
 : CURRENT FILING DATE: 2002-02-21
 : NUMBER OF SEQ. ID NOS: 47374

SEQ. ID NO 4614

LENGTH: 506

TYPE: PRT

ORGANISM: Burkholderia cepacia

US-60-360-039-4614

Query Match: 70.9%; Score 1894; DB 27; Length 506;

Best Local Similarity 70.6%; Pred. No. 1,30-186;

Matches 359; Conservative 62; Mismatches 82; Indels 2; Gaps 1;

QY 17 FPLKIRARYDNFTGTFWVAFAIGFYVQNLTPVTGQLLCEVASSGKRKIDLDLDAIKVVD 76

Db 13 FP--YKKQVANTIGGEWVAPADGEYVQNLTPVTGQLLCEVASSGKRKIDLDLDAIKVVD 70

QY 77 KWAHTSVQDRAALFKIADRMFQNLLELTAETWNGKPIRETSAALVPIADHF 146

Db 71 AKWKTSAADPANIIPNIAIPMFANIQPIAVATTLNKGKPIRETIAADIPDLADHFYFAG 140

QY 137 CIRAQEGGISEVDESETVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 196

Db 131 AVRAQEGGISEVDESETVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 190

QY 197 PLTPSVLLMEIVGILLPPGVVNVVNGAGVIGEYIATSKRIAKVAFSTGVG 256

Db 191 EOTPASILVVELIQLLPPGVVNVVNGFLEAGKPLASSPRIAKIATGETTIG 250

QY 257 YATONTIPVTELEGGKSPNIVFAVMDEEDAFDALKLEGFALFAFNGVCVCS 416

Db 251 YASONTIPVTELEGGKSPNIFADVMNEDUSEFDKALEGFAMFALNAGEVTTCTSKVLD 410

QY 317 ESTIYEFMEPAIRPVESTIRSGNPLDSVTOMGAQVSHGOLETIILNYIDIGKKEGAD 476

Db 311 EKLYDFMERALKRVAEIVQGDPLDPATMICAQASSPQLEKILSYIDIGHQEGAE 470

QY 377 RKLLGELEKDCYYLEPTILFGQNNMRVFOEEIFGPVLAVTTFTKMEALELAND 436

Db 371 ERNALGGELSKGYVYKPTVFRGNHNMRIFOEEIFGPVVSVTTFKDDDEALSAND 430

QY 447 GAVWTRCNRAYRFGRAIQAGRWNTNCHYATPAHAAFGYKQSGIGKRTHKMML 496

Db 431 GAGWTRCNRAYRFGRAIQAGRWNTNCHYATPAHAAFGYKQSGIGKRTHKMML 490

QY 497 TKCLLVSYSDKPLGLF 512

Db 491 TKNLVSYSDKPLGLFF 506

RESULT 13

US-60-360-039-7471
 : Sequence 7471, Application US/60360049
 : GENERAL INFORMATION:
 : APPLICANT: Cao, Yonqwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : FILE REFERENCE: 38-10(52052)A
 : CURRENT APPLICATION NUMBER: US/60/360,039
 : CURRENT FILING DATE: 2002-02-21
 : NUMBER OF SEQ. ID NOS: 47374

SEQ. ID NO 7471

LENGTH: 506

TYPE: PRT

ORGANISM: Burkholderia cepacia

US-60-360-039-7471

Query Match: 70.9%; Score 1894; DB 27; Length 506;

Best Local Similarity 70.6%; Pred. No. 1,30-186;

Matches 359; Conservative 62; Mismatches 82; Indels 2; Gaps 1;

QY 17 FPLKIRARYDNFTGTFWVAFAIGFYVQNLTPVTGQLLCEVASSGKRKIDLDLDAIKVVD 76

Db 13 FP--YKKQVANTIGGEWVAPADGEYVQNLTPVTGQLLCEVASSGKRKIDLDLDAIKVVD 70

QY 77 KWAHTSVQDRAALFKIADRMFQNLLELTAETWNGKPIRETSAALVPIADHF 146

Db 71 AKWKTSAADPANIIPNIAIPMFANIQPIAVATTLNKGKPIRETIAADIPDLADHFYFAG 140

QY 137 CIRAQEGGISEVDESETVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 196

Db 131 AVRAQEGGISEVDESETVAYHEHEPLGVVGGQIIPWFFLLMAWKMAPALAAAGNCV 190

QY 197 PLTPSVLLMEIVGILLPPGVVNVVNGAGVIGEYIATSKRIAKVAFSTGVG 256

Db 191 EOTPASILVVELIQLLPPGVVNVVNGFLEAGKPLASSPRIAKIATGETTIG 250

QY 257 YATONTIPVTELEGGKSPNIVFAVMDEEDAFDALKLEGFALFAFNGVCVCS 416

Db 251 YASONTIPVTELEGGKSPNIFADVMNEDUSEFDKALEGFAMFALNAGEVTTCTSKVLD 410

QY 317 ESTIYEFMEPAIRPVESTIRSGNPLDSVTOMGAQVSHGOLETIILNYIDIGKKEGAD 476

Db 311 EKLYDFMERALKRVAEIVQGDPLDPATMICAQASSPQLEKILSYIDIGHQEGAE 470

QY 377 RKLLGELEKDCYYLEPTILFGQNNMRVFOEEIFGPVLAVTTFTKMEALELAND 436

Db 371 ERNALGGELSKGYVYKPTVFRGNHNMRIFOEEIFGPVVSVTTFKDDDEALSAND 430

QY 447 GAVWTRCNRAYRFGRAIQAGRWNTNCHYATPAHAAFGYKQSGIGKRTHKMML 496

Db 431 GAGWTRCNRAYRFGRAIQAGRWNTNCHYATPAHAAFGYKQSGIGKRTHKMML 490

QY 497 TKCLLVSYSDKPLGLF 512

Db 491 TKNLVSYSDKPLGLFF 506

QY 303 QGEVTCPSRALVOESIYEREMERAIIRRVESIRSGNPLDSVTQMGAQVSHGOLETILNYI 462
 Db 296 QGEVTCPSRALVOESIYEREMERAIIRRVESIRSGNPLDSVTQMGAQVSHGOLETILNYI 462
 QY 463 DIGKKPGADVITGGPKKLIPEELKGGYILEPTILFQNNMRVFPQEEIFGPVLAVTIFRTM 422
 Db 456 DIGKQGAEVLIGGERNTFGDLGGYVVKPTVFKGHNMRIQEEIFGPVVSVTFTTUD 415
 QY 423 EEALELANDTOYGLGAGVWSENGNLAYKMGRCIGQGRVWTCYHAYPAHAAPGGYKQSGI 482
 Db 416 AEALHIANDLYGLGA:VWTRC:NPAYPR:PA:QAGVWTCYHAYPAHAAPGGYKQSGI 475
 QY 483 GRETHKMMLEHYQOTKCLIVSYSDKPIGLF 512
 Db 476 GREHKKMLDHYQOTKMMLVSYSPKKLGEF 505

Search completed: June 24, 2003, 10:30:44
 Job time : 188.362 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:35 ; Search time 55.428 Seconds
(without alignments)
2375.712 Million cell updates/sec

Title: US-09-830-751-8

Perfect score: 2673

Sequence: 1 MTNPPSAQIKPGYGFPIK HYQTKTIVSVSWKPLDIF 512

Scoring table BLASTM62

Gapop 10 0, Gapext 0.5

Searched: 1171708 seqs, 257149365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/FCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2673	100	0	512	US-09-830-751-8
2	2668	99	8	542	US-10-282-122A-43103
3	2668	99	8	542	US-10-282-122A-43103
4	2589	96	9	512	US-10-282-122A-75409
5	2567	96	0	534	US-10-282-122A-56331
6	2560	95	8	546	US-10-417-886-9975
7	2054	74	8	456	US-10-282-122A-72955
8	1918	71	8	506	US-10-369-493-17072
9	1917	71	7	495	US-10-369-493-15459
10	1917	71	7	495	US-10-369-493-15827
11	1917	71	7	495	US-10-369-493-16210
12	1915	71	6	567	US-10-282-122A-50457
13	1905	71	3	505	US-10-369-493-12218
14	1894	70	9	506	US-10-282-122A-49350
15	1894	70	9	506	US-10-369-493-4614
16	1894	70	9	506	US-10-369-493-7371
17	1890	70	7	495	US-10-369-493-11572
18	1878	70	3	506	US-10-369-493-17140
19	1870	70	0	504	US-10-369-493-19271
20	1868	69	9	493	US-10-369-493-4608
21	1868	69	9	493	US-10-369-493-7305
22	1868	69	4	506	US-10-282-122A-44659
23	1856	69	4	495	US-10-369-493-17869
24	1853	69	3	508	US-10-282-122A-47513
25	1849	69	2	515	US-10-369-493-23469
26	1847	69	1	492	US-10-469-493-20615

27	1839	68	8	507	6	US-10-282-122A-61783	Sequence 61783, A
28	1827	68	4	503	6	US-10-282-122A-44658	Sequence 44658, A
29	1823	68	2	514	6	US-10-417-886-9975	Sequence 9975, Ap
30	1821	68	1	494	6	US-10-369-493-21741	Sequence 21741, A
31	1819	68	1	506	6	US-10-282-122A-68054	Sequence 68054, A
32	1817	68	0	503	6	US-10-282-122A-63180	Sequence 63180, A
33	1808	67	6	527	6	US-10-282-122A-51390	Sequence 51390, A
34	1802	67	4	506	6	US-10-282-122A-65886	Sequence 65886, A
35	1798	67	4	507	6	US-10-282-122A-62787	Sequence 62787, A
36	1794	67	4	507	6	US-10-282-122A-64382	Sequence 64382, A
37	1795	67	2	682	6	US-10-366-683-31897	Sequence 31897, A
38	1795	67	2	682	6	US-10-419-128-31897	Sequence 31897, A
39	1791	67	0	506	6	US-10-282-122A-65344	Sequence 65344, A
40	1791	67	0	506	6	US-10-282-122A-77335	Sequence 77335, A
41	1783	66	7	506	6	US-10-369-493-8796	Sequence 8796, Ap
42	1773	66	3	506	6	US-10-369-493-14000	Sequence 14000, A
43	1770	66	2	507	6	US-10-156-761-14706	Sequence 14706, A
44	1759	65	8	534	6	US-10-446-203-10148	Sequence 10148, A
45	1758	65	8	506	6	US-10-282-122A-60172	Sequence 60172, A

ALIGNMENTS

RESULT 1

US-09-830-751-8

; Sequence 8, Application US/09830751

; GENERAL INFORMATION:

; APPLICANT: Suthers, Patrick F.

; TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant

; TITLE OF INVENTION: Organisms

; FILE REFERENCE: 960296, 96617

; CURRENT APPLICATION NUMBER: US/09/830,751

; CURRENT FILING DATE: 2000-08-30

; PRIOR APPLICATION NUMBER: 65/151,440

; PRIOR FILING DATE: 1999-08-30

; PRIOR APPLICATION NUMBER: PCT/US00/23878

; PRIOR FILING DATE: 2000-08-30

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO: 8

; LENGTH: 512

; TYPE: PRT

; ORGANISM: Escherichia coli

; US-09-830-751-8

Query Match: 100.0%; Score: 2674; DB: 5; Length: 512;
Best Local Similarity: 100.0%; Prod. No.: 1.76-236;
Matches: 512; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY	1	MTNPPSAQIKPGYGFPIKAKYUNF	IGGEWVAFAKSEYQNLTPVIGQLICEVASSG	60
DB	1	MTNPPSAQIKPGYGFPIKAKYUNF	IGGEWVAFAKSEYQNLTPVIGQLICEVASSG	60
QY	61	KRDTDLALDAHKYKWAHTSVQDPAALIFK	IADMEQNLLELLAIAFWNDKPIPEFS	120
DB	61	KRDTDLALDAHKYKWAHTSVQDPAALIFK	IADMEQNLLELLAIAFWNDKPIPEFS	120
QY	121	AADYPLALDHPYFASCI	RAQEGGISVDSSETVAYHFHPLGVVGQIIPWNPFLMASWK	180
DB	121	AADYPLALDHPYFASCI	RAQEGGISVDSSETVAYHFHPLGVVGQIIPWNPFLMASWK	180
QY	181	MAPALAAAGNCVVLKPAFTPLSLVLLMEI	VERLLPFGVNVVWNGAGVIDEYLATSKPIA	240
DB	181	MAPALAAAGNCVVLKPAFTPLSLVLLMEI	VERLLPFGVNVVWNGAGVIDEYLATSKPIA	240
QY	241	KVAFIGSTEVSQIQIMQVATQNIIPVTLEL	GKRSINIVFAVMDEEAFKALFPALEA	300
DB	241	KVAFIGSTEVSQIQIMQVATQNIIPVTLEL	GKRSINIVFAVMDEEAFKALFPALEA	300
QY	301	FNAGSEVCTSPALVHSEIVRFMEKPAI	PAVESIRSDNPILSVLQWACVSHQIETILN	450
DB	301	FNAGSEVCTSPALVHSEIVRFMEKPAI	PAVESIRSDNPILSVLQWACVSHQIETILN	450

Db 301 FNOGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETILN 360
 QY 361 YIDIGKKEGADVLTGGRRKLLFGLKDGYYLEPTILFGONNMRVFEELFGPVLAVTTFK 420
 Db 361 YIDIGKKEGADVLTGGRRKLLFGLKDGYYLEPTILFGONNMRVFEELFGPVLAVTTFK 420
 QY 421 TMEALELANTDYGICAGVWSRNGNLAYKMGRTGAGRWTCNCYHAYPAHAAGFGYKOS 480
 Db 421 TMEALELANTDYGICAGVWSRNGNLAYKMGRTGAGRWTCNCYHAYPAHAAGFGYKOS 480
 QY 481 GIGRETHKMMLEHYQOTKLLVSYSDDKPLGLF 512
 Db 481 GIGRETHKMMLEHYQOTKLLVSYSDDKPLGLF 512

RESULT 2

US-10-282-122A-43103
 ; Sequence 43103, Application US/10282122A

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Gidon
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, K.
 ; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 43103

; LENGTH: 542

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-282-122A-43103

Query Match

Best Local Similarity 99.8%; Score 2668; DB 6; Length 542;

Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPCEYGFPLKARYDNFISGHWAPADCEYYQNLTPVTGQLLCEVASSG 60
 Db 31 MTNNPSSAQIKPCEYGFPLKARYDNFISGHWAPADCEYYQNLTPVTGQLLCEVASSG 90
 QY 61 KRDIIDLALDAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 120

Db 91 KRDIIDLALDAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVISETVAYVHFHFPLGVCVQIIPWNPFLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIRAQEGGISEVISETVAYVHFHFPLGVCVQIIPWNPFLMASWK 210
 QY 181 MAVALAAGNCVVLKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 240
 Db 211 MAPALAAGNCVVLKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 270
 QY 241 KVAFTGSTEVGGQIMQYATONIIPTVLEAGGKSPNIVFADVMDEEDAFKALFGFALFA 400
 Db 271 KVAFTGSTEVGGQIMQYATONIIPTVLEAGGKSPNIVFADVMDEEDAFKALFGFALFA 440
 QY 301 FNOGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETILN 460
 Db 331 FNOGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETILN 490
 QY 361 YIDIGKKEGADVLTGGRRKLLFGLKDGYYLEPTILFGONNMRVFEELFGPVLAVTTFK 420
 Db 391 YIDIGKKEGADVLTGGRRKLLFGLKDGYYLEPTILFGONNMRVFEELFGPVLAVTTFK 450
 QY 421 TMEALELANTDYGICAGVWSRNGNLAYKMGRTGAGRWTCNCYHAYPAHAAGFGYKOS 480
 Db 451 TMEALELANTDYGICAGVWSRNGNLAYKMGRTGAGRWTCNCYHAYPAHAAGFGYKOS 510
 QY 481 GIGRETHKMMLEHYQOTKLLVSYSDDKPLGLF 512
 Db 511 GIGRETHKMMLEHYQOTKLLVSYSDDKPLGLF 542

RESULT 3

US 10 469 493 24602

; Sequence 24602, Application US/10 469493

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/460,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 24602

; LENGTH: 542

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-469-493-24602

Query Match

Best Local Similarity 99.8%; Score 2668; DB 6; Length 542;

Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPCEYGFPLKARYDNFISGHWAPADCEYYQNLTPVTGQLLCEVASSG 60
 Db 31 MTNNPSSAQIKPCEYGFPLKARYDNFISGHWAPADCEYYQNLTPVTGQLLCEVASSG 90
 QY 61 KRDIIDLALDAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 120
 Db 91 KRDIIDLALDAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVISETVAYVHFHFPLGVCVQIIPWNPFLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIRAQEGGISEVISETVAYVHFHFPLGVCVQIIPWNPFLMASWK 210
 QY 181 MAPALAAGNCVVLKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 240

211	MAPALAA	GN	CV	VL	KP	AP	I	TP	SV	LL	ME	IV	ED	LL	PG	VW	VW	NG	AG	SG	V	GE	Y	L	A	T	S	K	R	I	A	270																										
241	KV	AF	TG	ST	EV	GQ	Q	I	M	O	V	A	T	O	N	I	P	V	T	E	L	G	K	S	P	N	I	V	F	A	D	V	M	E	D	A	F	D	K	A	L	E	G	F	A	L	F	A	300									
271	KV	AF	TG	ST	EV	GQ	Q	I	M	O	V	A	T	O	N	I	P	V	T	E	L	G	K	S	P	N	I	V	F	A	D	V	M	E	D	A	F	D	K	A	L	E	G	F	A	L	F	A	330									
301	FN	OG	EV	CT	C	P	S	R	A	L	V	Q	E	S	T	I	V	E	R	M	E	R	A	I	R	V	E	S	T	I	R	S	C	N	P	I	D	S	T	Q	M	G	A	V	S	H	Q	L	E	T	I	L	N	360				
331	FN	OG	EV	CT	C	P	S	R	A	L	V	Q	E	S	T	I	V	E	R	M	E	R	A	I	R	V	E	S	T	I	R	S	C	N	P	I	D	S	T	Q	M	G	A	V	S	H	Q	L	E	T	I	L	N	390				
361	Y	I	D	I	G	K	E	G	A	D	L	T	G	R	R	K	L	E	G	L	K	D	G	Y	L	E	P	T	I	L	L	G	Q	N	N	M	V	P	O	E	E	I	F	G	P	V	L	A	T	T	F	K	420					
391	Y	I	D	I	G	K	E	G	A	D	L	T	G	R	R	K	L	E	G	L	K	D	G	Y	L	E	P	T	I	L	L	G	Q	N	N	M	V	P	O	E	E	I	F	G	P	V	L	A	T	T	F	K	450					
421	T	M	E	A	L	E	L	A	N	D	I	O	T	G	L	G	A	G	V	W	S	R	N	G	N	L	A	Y	K	M	G	R	G	L	U	A	G	K	V	M	I	N	C	Y	H	A	Y	A	H	A	A	F	G	Y	K	U	S	480
451	T	M	E	A	L	E	L	A	N	D	I	O	T	G	L	G	A	G	V	W	S	R	N	G	N	L	A	Y	K	M	G	R	G	L	U	A	G	K	V	M	I	N	C	Y	H	A	Y	A	H	A	A	F	G	Y	K	U	S	510
481	G	I	G	R	E	T	H	K	M	L	E	H	Y	Q	O	T	K	C	L	L	V	S	S	D	K	P	L	G	L	F	512																											
511	G	I	G	R	E	T	H	K	M	L	E	H	Y	Q	O	T	K	C	L	L	V	S	S	D	K	P	L	G	L	F	542																											

RESULT 4

RES-0014
US-10-1282-122A-75409
Sequence 75409, Application US/1082122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlstedt, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
1 FILE REFERENCE: ELITRA.034A
2 CURRENT APPLICATION NUMBER: US/16/282,122A
3 CURRENT FILING DATE: 2003-02-20
4 PRIOR APPLICATION NUMBER: 60/191,078
5 PRIOR FILING DATE: 2000-03-21
6 PRIOR APPLICATION NUMBER: 60/206,848
7 PRIOR FILING DATE: 2000-05-23
8 PRIOR APPLICATION NUMBER: 60/207,727
9 PRIOR FILING DATE: 2000-05-26
10 PRIOR APPLICATION NUMBER: 60/230,335
11 PRIOR FILING DATE: 2000-09-06
12 PRIOR APPLICATION NUMBER: 60/230,347
13 PRIOR FILING DATE: 2000-09-09
14 PRIOR APPLICATION NUMBER: 60/242,578
15 PRIOR FILING DATE: 2000-10-23
16 PRIOR APPLICATION NUMBER: 60/253,625
17 PRIOR FILING DATE: 2000-11-27
18 PRIOR APPLICATION NUMBER: 60/257,931
19 PRIOR FILING DATE: 2000-12-22
20 PRIOR APPLICATION NUMBER: 60/267,636
21 PRIOR FILING DATE: 2001-02-04
22 PRIOR APPLICATION NUMBER: 60/269,408
23 PRIOR FILING DATE: 2001-02-16
24 Remaining Prior Application data removed - See File Wrapper or PAM.
25 NUMBER OF SEQ ID NOS: 78614
26 SOFTWARE: PatentIn version 3.1
27 SEQ ID NO: 75409
28 LENGTH: 512

```

US-10-282-132A-75409

Query Match

Best Local Similarity 95.7%, Pred. No. 8.8e-229;
Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

```
Y 1 MNNP+SAQIKGCEYCFPLKAKYDNF+IGSEWVAPALGFFYYUNL+IPVTG+LLCEVASSG 60
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
D 1 MNNP+STRIGSEYCYPLKAKYDNF+IGTGWVAPALGFFYYUNL+IPVTG+GPLFEVASSG 60
   |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Y 61 KPDI+DALDAAHKVKVNAHTSVQDPAAL+LFK+IADRMF+UNL+FLI+LATAFTW+DNGKPI+PPTS 120
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 61 KKDI+DALDAAHKAKDKWAHTSVQDRAALL+LFK+IADRM+EQNLELL+LATAFTW+DNGKPI+RETS 120
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 121 AADV+PLALDH+EAYFAS+LRAPG+GI+SEVDS+ETVAYH+HEPL+GVVG+QI+IPWNP+PL+IMASW+K 180
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 121 AADI+PLADH+FRYFAS+LRAG+EGGI+SEVDS+ETVAYH+HEPL+GVVG+QI+IPWNP+PL+IMASW+K 180
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 181 MAPALA+AGNCV+LKPAR+TPLSV+LLME+IVGDL+LPGVVNVN+VGAGGV+IGEY+LATSK+RIA 240
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 181 MAPALA+AGNCV+LKPAR+TPLSV+LLME+IVGTL+PPGVVNVN+VGAGGV+IGEY+LATSK+RIA 240
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 241 KVAFT+GSTVG+QOI+MOYAT+ONI+IPVTL+ELGSK+SPN+IVFADV+MEDEAD+FDKALEG+FALFA 300
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 241 KVAFT+GSTVG+QOI+MOYAT+ONI+IPVTL+ELGSK+SPN+IFAI+VMDEE+AD+FDKALEG+FALFA 300
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 301 FNGQ+EWTC+PSPALQ+VEST+YRPF+MERAT+IPRV+EST+RS+GNPL+DSVT+QMGAVSH+QI+LETT+IN 360
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 301 FNG+EWTC+PSPALQ+VES+IYEP+PFMER+ALP+AVES+IP+SNPL+DS+ITLM+AVSH+QI+LETT+IN 360
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 361 YTDIG+KKGE+ADVL+IGSP+RKLI+EGEL+KGGY+ILF+PL+IFG+NNMK+VFGEE+IFG+PVLAV+TTFK 420
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 361 YLDIG+KKGE+ADVL+IGSP+RKEL+IGEL+KGGY+ILF+PL+IFG+NNMK+VFG+FF+IFG+PVLAV+TTFK 420
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 421 IMBE+ALELAND+TQYGLG+AVSW+SKN+INLAY+KMG+SLQAG+KVW+IN+CYHAY+FAHA+AF+SY+KQS 480
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 421 IMK+ALELAND+TQYGLG+AVSW+SPN+INLAY+KMG+SLQAG+KVW+IN+CYHAY+FAHA+AF+SY+KQS 480
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Y 481 GIGRE+THKMMLEHY+QOTK+CLLV+SYSDK+PLGLF 512
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
D 481 GIGRE+THKMMLEHY+QOTK+CLLV+SYSDK+PLGLF 512
   |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
```

RESULT 5

```

RES011 3
US:10-282-122A-56331
1 Sequence 56331, Application US/10-282-122A
2 GENERAL INFORMATION:
3 APPLICANT: Wang, Liangsu
4 APPLICANT: Zamudio, Carlos
5 APPLICANT: Malone, Cheryl
6 APPLICANT: Haselbeck, Robert
7 APPLICANT: Ohlsen, Kari
8 APPLICANT: Zyskind, Judith
9 APPLICANT: Wall, Daniel
10 APPLICANT: Trawick, John
11 APPLICANT: Carr, Grant
12 APPLICANT: Yamamoto, Robert
13 APPLICANT: Forsyth, R.
14 APPLICANT: Xu, H.
15 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
16 FILE REFERENCE: ELITRA.034A
17 PRIORITY APPLICATION NUMBER: US/10/282,122A
18 CURRENT FILING DATE: 2003-02-20
19 PRIOR APPLICATION NUMBER: 60/131,078
20 PRIOR FILING DATE: 2000-03-21
21 PRIOR APPLICATION NUMBER: 60/206,848
22 PRIOR FILING DATE: 2000-05-23
23 PRIOR APPLICATION NUMBER: 60/207,727
24 PRIOR FILING DATE: 2000-05-26
25 PRIOR APPLICATION NUMBER: 60/230,335
26 PRIOR FILING DATE: 2000-09-06
27 PRIOR APPLICATION NUMBER: 60/230,347
28 PRIOR FILING DATE: 2000-09-09
29 PRIOR APPLICATION NUMBER: 60/242,578
30 PRIOR FILING DATE: 2000-10-23
31 PRIOR APPLICATION NUMBER: 60/253,625

```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/261,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56331
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-56331

Query Match          96.08; Score 2567; DB 6; Length 534;
Best Local Similarity 94.5%; Pred. No. 16,226;
Matches 484; Conservative 17; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTNPPSAOIKPGEGYGFPLKIKARYDNFTGGWVAFADGGEYQNTLPVTGQLLCEVASSG 60
DB 23 MTNPPSSRIQPGEGYGFPLKIKARYDNFTGGWVADVDGGEYQNTLPVTGQPLCEIASSG 82
QY 61 KRDIIDLALDAAHKVKOKWAHTSVQDRAAILFKIADRMEOQLLELATAETWNGKPIRETS 120
DB 83 KRDIIDLALDAAHKAKDKWGHTSVQDRAAILFKIADRMEOQLLELATAETWNGKPIRETM 142
QY 121 AADVPLAIDHFRYFASCIRAOEGGISEVDSIVAYHFHEPIGVVQGIIPWNPFLLMASWK 180
DB 143 AADVPLAIDHFRYFASCIRAOEGGISEVDSIVAYHFHEPIGVVQGIIPWNPFLLMASWK 202
QY 181 MAPALAGNCVILKPKARLPFLSVLLMEIVGDLPPGVNVVNGAGGVIGEYLATSKRIA 240
DB 203 MAPALAGNCVILKPKARLPFLSVLLMEIVGDLPPGVNVVNGAGGELGEYLATSKRIA 262
QY 241 KVAFTGSTVGQIMQYATUNIIIPVTLEGGKSPNIVFADVMDDEDAFDFKALEGFALFA 300
DB 263 KVAFTGSTVGQIMQYATUNIIIPVTLEGGKSPNIVFADVMDDEDAFDFKALEGFALFA 322
QY 301 FNOGEVCTPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 360
DB 323 FNOGEVCTPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 382
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
DB 383 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 442
QY 421 TMEALELANDTOYGLCAGVWSNGNLAAYKMRGCTQAGVWNTNCHAYPAHAAPFGYKQS 480
DB 443 TMEALELANDTOYGLCAGVWSNGNLAAYKMRGCTQAGVWNTNCHAYPAHAAPFGYKQS 502
QY 481 GIGRETHKMLHEHYQQTCKLLVSYSOKPLGLF 512
DB 503 GIGRETHKMLHEHYQQTCKLLVSYSOKPLGLF 534

RESULT 6
US-10-417-886-9975
; Sequence 9975, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10417,886
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/252,691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
```

```
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 9975
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-417-886-9975

Query Match          95.88; Score 2560; DB 6; Length 546;
Best Local Similarity 94.4%; Pred. No. 4,662,226;
Matches 484; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNPPSAOIKPGEGYGFPLKIKARYDNFTGGWVAFADGGEYQNTLPVTGQLLCEVASSG 60
DB 35 MTNPPSSRIQPGEGYGFPLKIKARYDNFTGGWVAFADGGEYQNTLPVTGQPLCEIASSG 94
QY 61 KRDIIDLALDAAHKVKOKWAHTSVQDRAAILFKIADRMEOQLLELATAETWNGKPIRETS 120
DB 95 KRDIIDLALDAAHKAKDKWGHTSVQDRAAILFKIADRMEOQLLELATAETWNGKPIRETM 154
QY 121 AADVPLAIDHFRYFASCIRAOEGGISEVDSIVAYHFHEPIGVVQGIIPWNPFLLMASWK 180
DB 155 AADVPLAIDHFRYFASCIRAOEGGISEVDSIVAYHFHEPIGVVQGIIPWNPFLLMASWK 214
QY 181 MAPALAGNCVILKPKARLPFLSVLLMEIVGDLPPGVNVVNGAGGVIGEYLATSKRIA 240
DB 215 MAPALAGNCVILKPKARLPFLSVLLMEIVGDLPPGVNVVNGAGGELGEYLATSKRIA 274
QY 241 KVAFTGSTVGQIMQYATUNIIIPVTLEGGKSPNIVFADVMDDEDAFDFKALEGFALFA 300
DB 275 KVAFTGSTVGQIMQYATUNIIIPVTLEGGKSPNIVFADVMDDEDAFDFKALEGFALFA 344
QY 301 FNOGEVCTPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 360
DB 335 FNOGEVCTPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 394
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
DB 395 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 454
QY 421 TMEALELANDTOYGLCAGVWSNGNLAAYKMRGCTQAGVWNTNCHAYPAHAAPFGYKQS 480
DB 455 TMEALELANDTOYGLCAGVWSNGNLAAYKMRGCTQAGVWNTNCHAYPAHAAPFGYKQS 514
QY 481 GIGRETHKMLHEHYQQTCKLLVSYSOKPLGLF 512
DB 515 GIGRETHKMLHEHYQQTCKLLVSYSOKPLGLF 546

RESULT 7
US-10-282-122A-72955
; Sequence 72955, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianqun
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: FLIIPA 934A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; FEIOP APPLICATION NUMBER: 60/220,948
; FEIOP FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,727
```


; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15459
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15459

Query Match 71.7%; Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9%; Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALAAHVKVUKWAHT 81
DB 5 KPRYDNFICGAWAPQGGYFNITPTTKVFTSTAKSTAPDIEAALDAHAHAAKAWGKT 64
QY 82 SVQDRAAILFKIADRMFQNIETATAPTWNCKPIRETSAADVPLAIDHFRYFASCIRAQ 141
DB 65 STTDRSNVLKIAIDRIQNLLELLAYAETWDCNKPVRETINADVPLCVDFHFRYFACIRAQ 124
QY 142 EGGISEVDSETVAYHEHEPLGVVGGQIIPWNPFLLMASWMAKALAGNCVVLKPARLPL 201
DB 125 EGGISEIDSTIAYHHEPLGVVGGQIIPWNPFLLMACWKLAPALAGNCVVMKPAEQTPA 184
QY 202 SVLLMFIVGDLPPGVVNVVNGAGVIGEYLATSKRIAKVAFTGSTEVGGQIMOVATON 261
DB 185 SILVLMFIVGDLPPGVVNVVNGFLEAGKPLASNPRIAKIAFTGETTIGRLIMOVASON 244
QY 262 IIPVTELGKSPNIVFADVMDEDAFFDKALGFALFAPNGEVCTCPSRALVOESIYE 321
DB 245 LIPVTELGKSPNIFADVMDEDDFLDKAVRGFVLFAFNGEVCTCPSRALVOESIYE 304
QY 322 RFMEKALRRVESIRSGNPLDSVTOMCAQVSHGOLETLINXIDIGKKGADVLTCGRKLL 381
DB 305 TFMEKALRRVAAIKQGNPLDPNTWGAQASSGOLEKILSYIDIGKQGAELDIGGERNML 364
QY 382 EGELKGGYLLPTILFGQNNMRVFOEIEFCPLAVITTEKTMEEALELANDTQYGLGAGW 441
DB 365 DGELSDGFYVKPTVFKGNKMRVFOEIEFCPLAVITTEKTMEEALELANDTQYGLGAGW 424
QY 442 SRNCLAYKMGKGIQAGRVWTCYHAYPAHAAGGYKQSGIGRETHKMMLEHYQOTKLL 501
DB 425 SRASRLYRMGRIAGRVWTCYHAYPAHAAGGYKQSGIGRETHKMMLEHYQOTKLL 484
QY 502 VSYSDKPLGLF 512
DB 485 VSYSPKALGFF 495
  
```

RESULT 10
 US-10-369-493-15827
 ; Sequence 15827, Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15827
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15827

Query Match 71.7%; Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9%; Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALAAHVKVUKWAHT 81
DB 5 KPRYDNFICGAWAPQGGYFNITPTTKVFTSTAKSTAPDIEAALDAHAHAAKAWGKT 64
QY 82 SVQDRAAILFKIADRMFQNIETATAPTWNCKPIRETSAADVPLAIDHFRYFASCIRAQ 141
DB 65 STTDRSNVLKIAIDRIQNLLELLAYAETWDCNKPVRETINADVPLCVDFHFRYFACIRAQ 124
QY 142 EGGISEVDSETVAYHEHEPLGVVGGQIIPWNPFLLMASWMAKALAGNCVVLKPARLPL 201
DB 125 EGGISEIDSTIAYHHEPLGVVGGQIIPWNPFLLMACWKLAPALAGNCVVMKPAEQTPA 184
QY 202 SVLLMFIVGDLPPGVVNVVNGAGVIGEYLATSKRIAKVAFTGSTEVGGQIMOVATON 261
DB 185 SILVLMFIVGDLPPGVVNVVNGFLEAGKPLASNPRIAKIAFTGETTIGRLIMOVASON 244
QY 262 IIPVTELGKSPNIVFADVMDEDAFFDKALGFALFAPNGEVCTCPSRALVOESIYE 321
DB 245 LIPVTELGKSPNIFADVMDEDDFLDKAVRGFVLFAFNGEVCTCPSRALVOESIYE 304
QY 322 RFMEKALRRVESIRSGNPLDSVTOMCAQVSHGOLETLINXIDIGKKGADVLTCGRKLL 381
DB 305 TFMEKALRRVAAIKQGNPLDPNTWGAQASSGOLEKILSYIDIGKQGAELDIGGERNML 364
QY 382 EGELKGGYLLPTILFGQNNMRVFOEIEFCPLAVITTEKTMEEALELANDTQYGLGAGW 441
DB 365 DGELSDGFYVKPTVFKGNKMRVFOEIEFCPLAVITTEKTMEEALELANDTQYGLGAGW 424
QY 442 SRNCLAYKMGKGIQAGRVWTCYHAYPAHAAGGYKQSGIGRETHKMMLEHYQOTKLL 501
DB 425 SRASRLYRMGRIAGRVWTCYHAYPAHAAGGYKQSGIGRETHKMMLEHYQOTKLL 484
QY 502 VSYSDKPLGLF 512
DB 485 VSYSPKALGFF 495
  
```

RESULT 11
 US-10-369-493-16210
 ; Sequence 16210, Application US/10469494
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10469494
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 16210
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-16210

Query Match 71.7%; Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9%; Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALAAHVKVUKWAHT 81
DB 5 KPRYDNFICGAWAPQGGYFNITPTTKVFTSTAKSTAPDIEAALDAHAHAAKAWGKT 64
  
```

QY 82 SWQDRAALFKIADMEQNLLELATAETWNGKEIKETSAOYPLAIDHRYFASQIRAO 141
Db 65 STTDSNVLKLIADRIEQLNLELLAYATWNGKPVRETTNADVPICVDHPRYPAGAIRAO 124
QY 142 EGGISEVRSFTVAYHREPLGVVGGQIIPWNPPLIMASWKNAPALAAQNVVLIKPPAPLTPL 201
Db 125 EGGISEIDSDTIAYHREPLGVVGGQIIPWNPPLMACWKLAPALAAQNCVVMKPAQOTPA 184
QY 202 SVLLIMEIVGDLPLPGVNVVNGAGCVIGEYLATSKRIAKVAFTGTSTEVGQIQIMQATON 261
Db 185 SILVIMEIVGDLPLPGVNVVNGFGLIAGKPLASNPRIAKIAFTGTITGRIMQVASON 244
QY 262 IIPVTELECGKSPNIVADVNDDEDAFFDKALCEGALFAFNOCGEVCTCTSRALVORSIYE 321
Db 245 LIPVTELECGKSPNIFADVNAELGALFIDRAVESEVLFLENQSEVCTCTSRALVORSIYE 304
QY 322 RMEHAIIRVSEIFSGNPLDSVTOMGACVSHGOLETILNYIDIGKREGADVLPGRKELL 381
Db 305 TMEKALKRVAIKGNPLDNTMVGACASSEQLKILSYIDIGKJGJAEILIGGERNML 364
QY 382 EGELKDGYYLEPTILFGONNMNVFOEEIFGVPVLAFTTEKMEBALELANDTOYGLGAGVW 441
Db 365 DEGLSDGVPVFTVKHKNKRVFOEEIFGVPVSVITPKDEADALAIANDTLVGLGAGVW 424
QY 442 SRNGLAYKMGRIQAGRVWNTNCHAYPAHAAGGKYGKSGIGRETHKMMLEHYQOTKCLL 501
Db 425 SRDASRLVPMGRALQAGRVWNTNCHAYPAHAAGGKYGKSGIGRETHKMMLEHYQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 12

US-10-282-122A-50457

; Sequence 50457, Application US/10282122A

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ghisla, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003 02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/254,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ IN NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50457
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50457

Query Match

71.6%; Score 1915; DB 6; Length 567;

Best Local Similarity 72.0%; Pred. No. 9,1e-167;

Matches 357; Conservative 60; Mismatches 77; Indels 2; Gaps 1;

QY 17 FPLKLRARYDNFICGENVAVPADGEYYQNLTPVIGQLLCEVASSCKRDLIDALDAHVKVD 76
Db 74 FP--YRKYGFICGENVAVPVGGYFDNVSPVTRPPTAIPRSREADIELADAAHAAKA 131
QY 77 KWAHTSVQDCAALFKIADPMENQLELIATAFTWNGKPTFTTSAFVPIATIHPPYFAS 136
Db 132 GWAAGGAEPANVLLPIADPMENI/TPLAVAFITDNCCKPLPTTAAADVPLAIDHPPYFAG 191
QY 137 CIRACPGSTISVDSFTVAYHHPPLGVVGQIIPWNPPLIMASWKNAPALAAQNCVVLKPA 196
Db 192 CIRAQEGSIADIGDMVAYHHPPLGVVGQIIPWNPPLMAANKLAPALAAQNCVVLKPA 251
QY 147 PLTPLSVLLIMEIVGDLPLPGVNVVNGAGCVIGEYLATSKRIAKVAFTGTSTEVGQIMQ 256
Db 252 EQTFASILVFAELIQDLPLPGVNVVNGAGCVIGEYLATSKRIAKVAFTGTSTEVGQIMQ 311
QY 257 YATONIIPTVTELECGKSPNIVADVNDDEDAFFDKALCEGALFAFNOCGEVCTCTSRALV 316
Db 312 YASENLIPVTELECGKSPNIFADVNDDEDAFFDKALCEGALFAFNOCGEVCTCTSRALV 371
QY 317 PSIVPRMEHAIIRVSEIFSGNPLDSVTOMGACVSHGOLETILNYIDIGKREGADVLPGR 376
Db 372 ESIYDFPIERALKRVFAIKGHPULSOLMIGACASAPVQKILSYIDIGKJGJAEILIGGER 431
QY 377 RKILIFGRKIKSYVLEPTILFGONNMNVFOEEIFGVPVLAFTTEKMEBALELANDTOY 436
Db 432 ERNVLGELAGSYVVKPTVFRGHKNKRVFOEEIFGVPVLAFTTEKMEBALELANDTOY 491
QY 437 GAGVWSRNGSLAYKMGRIQAGRVWNTNCHAYPAHAAGGKYGKSGIGRETHKMMLEHY 496
Db 492 GAGVWTRNGSLAYKMGRIQAGRVWNTNCHAYPAHAAGGKYGKSGIGRETHKMMLEHY 551
QY 497 TKCLLVSYSDKPLGLF 512
Db 552 TKNLLVSYSEKPLGFF 567

RESULT 13

US 10 369-493-12218

; Sequence 493, Application US/10369493

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52953)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003 02-28
; PRIOR APPLICATION NUMBER: US 60/160,039
; PRIOR FILING DATE: 2002 02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12218
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12218

```

Query Match          71.4%  Score 1905;  DB 6;  Length 505;
Best Local Similarity 70.3%;  Pred No 6, 20-166;
Matches 352;  Conservative 63;  Mismatches 84;  Indels 2;  Gaps 1;

QY 14 EYGPFLK--LKARYDNF IGGEWVAPALGYYQNLTPVTGQLLEVASSGKRDIDLALDAK 71
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 5 EFSRVPKAPDKRYGNF IGKWKTFPSRFFENHSPVNGLLCEVARSDADDIEALDA 64
QY 72 HKYDKKAHSTSVODRAAILFKIADRMQNLIELLAETWNGKPIPETSAADVPLAIDHF 131
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 65 HAAKDAGRTSVASRSLINRIADRMEENLDLLACAEWDNGKPIRETIVADPLAIDHF 124
QY 132 RYFASCIARAGGGTSEVDSSTVAYHEPGLVGVGQIIPWFPPLMASKMAPALAAAGNCV 191
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 125 RYFASVYRGGBSLSQIDDDTVAYHEPGLVGVGQIIPWFPPLMACWKAPALAAAGNCV 184
QY 192 VLKPARITPTSVLLIMFIVGKLLPPGVVNVVNGAGVIGFYLATSKPIAKVAFSTGTEVG 251
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 185 VLKPAEQTPAAILWADLIGDLPVGVNLVNGFLEAGKPLASSPRIAKTAFTGTTTG 244
QY 252 QOIMOVATONIIPVTELEGKSPNIVFADVMEDEADFUKALEGFALFANFNGEVCYTCPS 311
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 245 RLIMQYASQNLIPVTELEGKSPNIFKDVVAEDDDDFDKALEGFVFMALNNGEVCYTCPS 304
QY 312 RALVQESIYRPFMERALIRVESTIRSGNPILDSVTQMGAVSHGULETILNYIDIGKKEGAD 371
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 305 RALHESIYRPFMERALKVEAIVQGDPLDPATMIGCAOASSSQLEKILSYIDIGRQEGAE 364
QY 372 VLTGGRKLLLEGELKDCYLEPTILEGQNNMRVFOEELFGPVLAVTTFKTMEEALELAND 431
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 365 VLTGGARNVLPGLAGSYVYKPIFVRGHNMKRFQEEIFGPVSVVTTFKDDDEALSAND 424
QY 432 TQGLGAGVMSRNLAYKMGKGIQAGRVWNTNCYHAYPAHAAGFYKSGGIGRETHKMKML 491
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 425 TLYGLGAGVWTRGNRAYRFGRAIQAGRVWNTNCYHAYPAHAAGFYKSGGIGRETHKMKML 484
QY 492 EHYQOTKCLLVSYSDKPLGLF 512
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 485 DRYQOTKMLVSYSPKRLGPF 505

RESULT 14
US-10-282-122A-49350
: Sequence 44456; Application US/0282122A
: GENERAL INFORMATION:
: APPLICANT: Wang, Lianqun
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlson, Karl
: APPLICANT: Zyskind, Judith
: APPLICANT: Walli, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, K.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-23
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/240,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23

```

```

: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,941
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,408
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 49350
: LENGTH: 506
: TYPE: PRT
: ORGANISM: Burkholderia funiformis
US-10-282-122A-49350

Query Match          70.9%  Score 1894;  DB 6;  Length 506;
Best Local Similarity 70.6%;  Pred No 6, 40-166;
Matches 359;  Conservative 62;  Mismatches 82;  Indels 2;  Gaps 1;

QY 17 FPLKLRARYNFIPIGGWVAPALGYYQNLTPVTGQLLEVASSGKRDIDLALDAK 76
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 13 FP--YRKQYANFIPIGGWVAPALGYYQNLTPVTGQLLEVASSGKRDIDLALDAK 70
QY 77 KWAHLSVQDPAALIFKIALHMPUNLELEALALWLNKPKIKELSAADVPLAIDHRYFAS 136
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 71 AWKKTSAADKANILNKAIDRMANLQRLAVATLUNGKPLRETAAIDPLAIDHRYFAG 150
QY 137 QTHAQGRGTSFVTSPTVAYHEPGLVGVGQIIPWFPPLMASKMAPALAAAGNCV 196
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 131 AVRAQEGSLSEIHDHDTVAYHEPGLVGVGQIIPWFPPLMASKMAPALAAAGNCV 190
QY 197 RLTPLSVLLIMELVGDILPPGVVNVVNGAGVIGFYLATSKPIAKVAFSTGTEVGQIMQ 256
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 191 EQTPASITLVLTQLDPPGVNLVNGFLEAGKPLASSKRIAKTAFTGTTTGELDLMQ 250
QY 257 YATQNIIPVTELEGKSPNIVFADVMEDEADFUKALEGFALFANFNGEVCYTCPSRAVQ 316
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 251 YASQNIIPVTELEGKSPNIFADVMEDEADFUKALEGFALFANFNGEVCYTCPSRAVQ 310
QY 317 ESIYRPFMERALIRVESTIRSGNPILDSVTQMGAVSHGULETILNYIDIGKKEGAD 376
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 311 ESIYRPFMERALIRVESTIRSGNPILDSVTQMGAVSHGULETILNYIDIGKKEGAD 370
QY 377 RKKLEGLKDCGYLEPTILEGQNNMRVFOEELFGPVLAVTTFKTMEEALELAND 436
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 371 ERNALGGLSKGYVYKPIFVRGHNMKRFQEEIFGPVSVVTTFKDDDEALSAND 430
QY 437 GAVVSPNGLIAYKMGKGIQAGRVWNTNCYHAYPAHAAGFYKSGGIGRETHKMKML 496
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 431 GAVVSPNGLIAYKMGKGIQAGRVWNTNCYHAYPAHAAGFYKSGGIGRETHKMKML 490
QY 497 TKCLLVSYSDKPLGLF 512
D 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
D 491 TKMLVSYSDKPLGLF 506

RESULT 15
US-10-369-493-4614
: Sequence 4614; Application US/10-069493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 48-10(52052)B
: CURRENT APPLICATION NUMBER: 08/137,469,493
: CURRENT FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/460,049

```

```

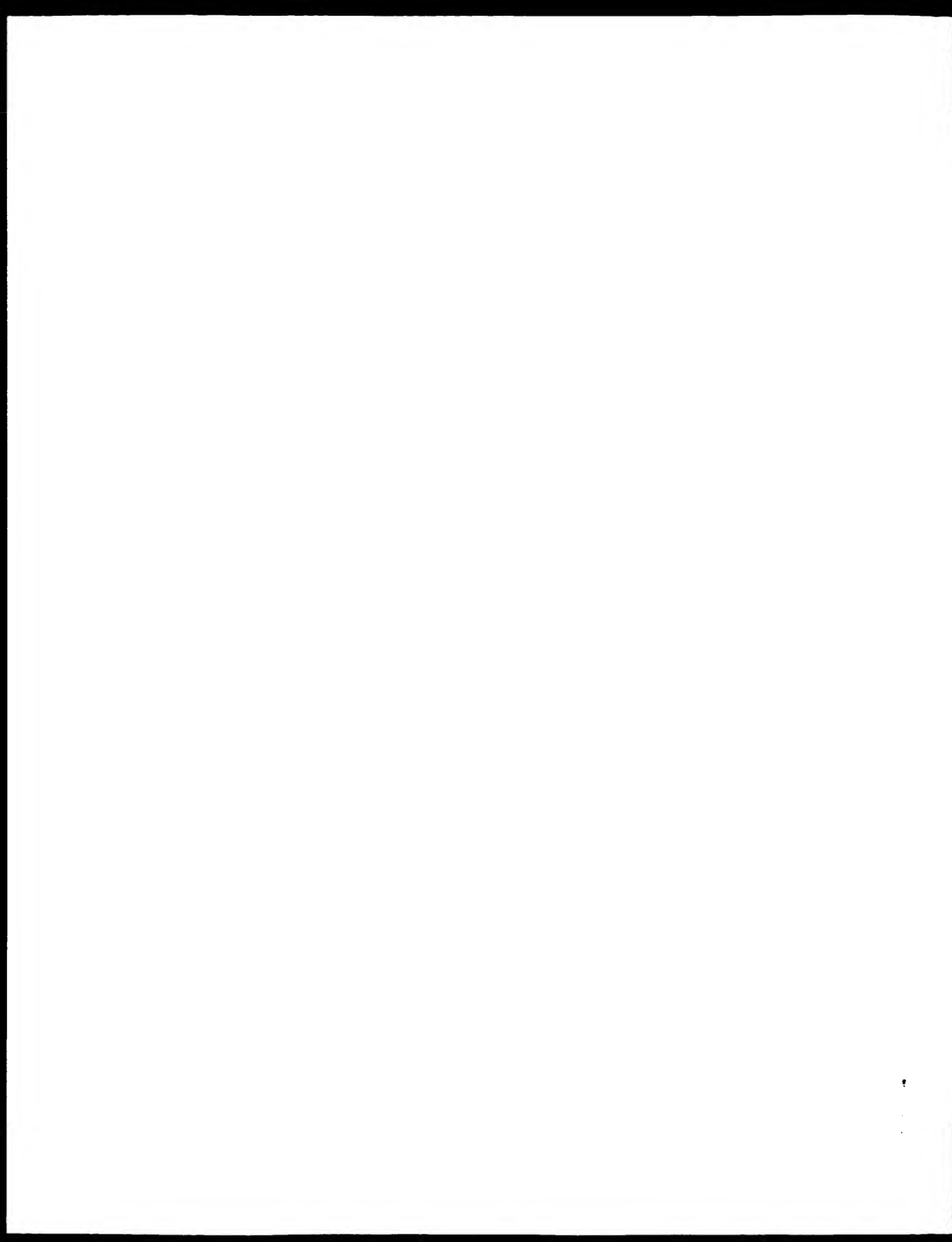
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4614
; LENGTH: 506
; TYPE: PR1
; ORGANISM: Burkholderia fungo
US-10-369-493-4614

```

Query Match 70.98; Score 1894; DB 6; Length 506;

QY	17	FPLKLBARYDNF	IGCEWVAPADGEYONLT	IPVTGQLLCEVASSGKKDD	LDLALDAAHKVKD	76	
Db	13	PE--YKKOANF	IGCEWVKVGGVEFUND	SPITCEAFTSIPRSEADVELAD	AAHRAKRT	70	
QY	77	KWAHTSVODRAA	ILFKLADRMEOINLELLATA	FAETWDNGKPIRETSAADVP	LAIIDHPRYFAS	136	
Db	71	ANGKTSAADRANILNR	TADMEANLOH	LAFAETIDNGKPI	PETTAADIPLAIDHPFYFAG	130	
QY	137	CLRAQEGGJISEVDS	ETVAYHIFHEPLGVVGU	IPWNEPFLMASMKMAPALA	AGNCVVLKPA	196	
Db	131	AVRAQEGGSLSEIDHUT	VAYHIFHEPLGVVGU	IPWNEPFLMAVWKALAPALA	AGNCVVLKPA	190	
QY	197	RLTPLSLVLLMEI	VGDLILPGCVNVNNGAGSV	IGEYLATSKRIAKVAFTG	STEVGGQIMQ	256	
Db	191	EOTPASILVLVEL	IODLLPGCVLNVNNGSL	EACKPLASKRIAKTAFT	TGTTTGRLLMQ	250	
QY	257	YATONIPVTL	ELGCKSPNIVFALVIMPE	DEDAFFDKALEGFALFAT	NOGEVCTCPSKRALVQ	316	
Db	251	VASQNIIPVTL	ELGCKSPNIFFAFDMN	EDDSFFDKALEGFAMFAL	NOGEVCTCPSRVLID	310	
QY	317	ESIVPERFEMERAL	IPVESIRSGNPLD	SVTOMGAQVSHQLETL	ILNYIDTGKKRCADVLGG	376	
Db	311	EKIYPERFEMERAL	KVVAALTCGHP	LDTKMTIGAGASUBLEK	ILSVYDLGKGQESACELGG	370	
QY	377	RKLELGELKDG	GYLYLEPTILF	LGONNMRFVEEITFG	VPVLAFTTFTKTMEEALELAND	TOYCL	436
Db	371	EFNAUGGFI	SKGYGVKPTVFP	GNKKMP	IFOEELFGPVSVTTFRNEEEALE	LANDTLYCL	430
QY	437	GAGVWSRNGN	ILAYKMGRCIGAG	KVWTCNCHAYFAHAA	FGSYKUSGTIGRTHKMMLEHYQ	496	
Db	431	GAGVWTRGCT	PAYRFGQIGAG	RWTCNCHAYFAHAA	FGSYKUSGTIGRHNHMM	LDHYQ	490
QY	497	TKCLLVSYSDK	PLGLF	512			
Db	491	TKNLLVSYSDK	PLGFF	506			

Search completed: June 24, 2003, 10:35:24
Job time : 57.428 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: Tue 24, 2003, 10:18:05, Search time: 19.944 seconds
(without alignments)
2512.114 Million cell updates/sec

Title: us-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQIKPGYGFPLK... HYGQTKCLVSYSDKPLGIF 512

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	99.8	542	1 S47809	aldehyde dehydrog
2	2651	99.2	542	2 H91186	aldehyde dehydrog
3	2651	99.2	542	3 C86033	aldehyde dehydrog
4	2589	96.9	512	2 AF0977	lactaldehyde dehyd
5	1918	71.8	506	2 R87603	aldehyde dehydrog
6	1910	71.5	506	1 A42597	aldehyde dehydrog
7	1895	70.9	505	2 AC2849	aldehyde dehydrog
8	1895	70.9	540	2 E97626	chloroacetaldehyde
9	1878	70.3	506	2 A93735	aldehyde dehydrog
10	1849.5	69.2	515	2 H75589	aldehyde dehydrog
11	1848.5	69.2	505	2 AF3470	aldehyde dehydrog
12	1802	67.4	506	2 F83142	aldehyde dehydrog
13	1798	67.3	507	2 F70827	probable aldehyde
14	1791	67.0	506	2 H83396	probable aldehyde
15	1791	67.0	506	2 F82452	probable aldehyde
16	1682	62.9	506	2 T09437	probable aldehyde
17	1682	62.9	541	2 A82276	aldehyde dehydrog
18	1085.5	40.6	495	2 C89778	hypothetical prote
19	1032	38.6	498	2 C83717	NADP-dependent ald
20	1010	37.8	501	2 T31265	aldehyde dehydrog
21	986	34.4	445	2 H66614	aldehyde dehydrog
22	981	36.7	519	1 I48966	aldehyde dehydrog
23	975	36.5	519	1 S03564	aldehyde dehydrog
24	971	36.3	517	1 DEH022	aldehyde dehydrog
25	970	36.3	503	2 T39216	aldehyde dehydrog
26	953	35.7	520	1 S09030	aldehyde dehydrog
27	952	35.6	500	1 S03364	aldehyde dehydrog
28	949	35.5	517	1 A40872	aldehyde dehydrog
29	940.5	35.2	519	1 S67286	probable aldehyde

30 940.5 35.2 520 1 S50576 Probable aldehyde
31 932 34.9 544 2 D88442 Protein F54D8.3 [i
32 931.5 34.8 497 1 A29055 aldehyde dehydrog
33 930 34.8 542 2 T02301 aldehyde dehydrog
34 928 34.7 496 2 C87638 aldehyde dehydrog
35 925.5 34.6 499 2 S74224 aldehyde dehydrog
36 918.5 34.4 549 2 T03983 rf2 nuclear restor
37 918 34.3 538 2 I06583 aldehyde dehydrog
38 913.5 34.2 490 2 C83675 glycine betaine al
39 913 34.2 519 2 C86372 hypothetical prote
40 910 34.0 490 2 A64629 glycine betaine al
41 905 33.9 512 1 A55684 aldehyde dehydrog
42 895 33.5 509 1 S14629 aldehyde dehydrog
43 891 33.3 497 2 B82981 Probable aldehyde
44 888.5 33.2 794 2 P95942 Probable aldehyde
45 886.5 33.2 494 2 B95411 Probable aldehyde

ALIGNMENTS

RESULT 1
S47809
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) aldR [similarity] - Escherichia coli (strai
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 03-Jun-2002
C:Accession: S47809; B57259; F65158
R:Plunkett, G.
Submitted to the EMBL Data Library, March 1994
A:Reference number: S47809
A:Accession: S47809
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <PLU>
A:Cross-references: EMBL: U00002, NID: g46f582 PDB: AAF18555 1, PDB: g912476
J. Xu, J. Johnson, R.C.
J. Bacteriol. 177, 3166-3175, 1995
A:Title: aldR, an RpoS dependent gene in Escherichia coli encoding an aldehyde dehydr
A:Reference number: A57259; MUID: 95286498; PMID: 7768815
A:Accession: B57259
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 31-542 <XUA>
A:Cross-references: GB: I40742; NID: g712824; PDB: AAC36930 1, PDB: g712825
R. Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Petina, N.T.; Burkand, V.; Riley, M.,
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID: 97426617; PMID: 9278503
A:Accession: F65158
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-542 <BLAT>
A:Cross-references: GB: AF000436, GB: U000026, NID: g2367246, PDB: AAT76622 1, PDB: g17500
A:Experimental source: strain P-12, substrain M31655
C:Genetics.
A:Gene: aldR
A:Start codon: GTG
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD, oxidoreductase
F: 93.357/Domain: aldehyde dehydrogenase homology; ALDE
F: 298.337/Active site: Glu, Cys #status predicted
F: 494/Binding site: NAD (Cys) #status predicted

Query Match 59.2%, 50.1%, 26.6%, 78.1%, 100%, 542;
Best Local Similarity 99.8%; Pred No. 8.5e-198;
Matches 511, Conservative 0, Mismatches 1, Indels 0, Gaps 0;

QY 1 MTNPPSAQIKPGYGFPLKIKAHYUNFDSSEWVAALHRYVUNLIPVIGQLACFEVASSIS 60
|||||
DB 31 MTNPPSAQIKPGYGFPLKIKAHYUNFDSSEWVAALHRYVUNLIPVIGQLACFEVASSIS 60
|||||
QY 61 KPETGLALDAHKKVKRWNIISVCTFAALFFKADHMEINIFILATAIWEINPKPIRETS 120

Db 91 KRDLALDAAHKVKWKWHTSVQDRAA LLFKLAIDMEQNLELLATAETWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 210
 QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 240
 Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 270
 QY 241 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 Db 271 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 330
 QY 301 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 360
 Db 331 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 390
 QY 361 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 420
 Db 391 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 450
 QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGIQAGRVWTCNRYHAYPAHAAGFGYKQS 480
 Db 451 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGIQAGRVWTCNRYHAYPAHAAGFGYKQS 510
 QY 481 GIGRETHKMMLEHYOQTKLLVSYSDKPLGLF 512
 Db 511 GIGRETHKMMLEHYOQTKLLVSYSDKPLGLF 542

RESULT 2

H91186
 aldehyde dehydrogenase B [imported] - Escherichia coli (strain O157:H7, substrain RMD 0
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug 2001
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and qenc
 A:Reference number: A99620; MIMD:21156231; PMID:11258795
 A:Accession: H91186
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-542 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAH37887.1; PID:q13363935; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 05-09952
 C:Genetics:
 A:Gene: Ecs4464
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 99.2%; Score 2651; DB 2; Length 542;
 Best Local Similarity 99.2%; Pred. No. 1.7e 196;
 Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 60
 Db 31 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 90
 QY 61 KRDLALDAAHKVKWKWHTSVQDRAA LLFKLAIDMEQNLELLATAETWNGKPIRETS 120
 Db 91 KRDLALDAAHKVKWKWHTSVQDRAA LLFKLAIDMEQNLELLATAETWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 210
 QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 240
 Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 270

QY 241 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 Db 271 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 330
 QY 301 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 360
 Db 331 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 390
 QY 361 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 420
 Db 391 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 450
 QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGIQAGRVWTCNRYHAYPAHAAGFGYKQS 480
 Db 451 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGIQAGRVWTCNRYHAYPAHAAGFGYKQS 510
 QY 481 GIGRETHKMMLEHYOQTKLLVSYSDKPLGLF 512
 Db 511 GIGRETHKMMLEHYOQTKLLVSYSDKPLGLF 542

RESULT 3

G86033
 aldehyde dehydrogenase B [similarity] - Escherichia coli (strain O157:H7, substrain E
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
 C:Accession: G86033
 R:Perna, N.T.; Plunkett, III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Kose, D.J.; May
 Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MIMD:21074935; PMID:11206551
 A:Accession: G86033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-542 <SFO>
 A:Cross-references: GB:AE005174; NID:q12518325; PIDN:AAG58731.1; GSPDB:GN00149; UMDP:
 A:Experimental source: strain O157:H7, substrain Ehu933
 C:Genetics:
 A:Gene: aldH
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 99.2%; Score 2651; DB 2; Length 542;
 Best Local Similarity 99.2%; Pred. No. 1.7e 196;
 Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 60
 Db 31 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 90
 QY 61 KRDLALDAAHKVKWKWHTSVQDRAA LLFKLAIDMEQNLELLATAETWNGKPIRETS 120
 Db 91 KRDLALDAAHKVKWKWHTSVQDRAA LLFKLAIDMEQNLELLATAETWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIQAEGGISEVDSETVAYHFHEPLGVGVGQIIPWNPFLLMASWK 210
 QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 240
 Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGIVIGEYLAISKRIA 270
 QY 241 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 Db 271 KVAFTGSTEVGQOIMQYATONIPVTLELGKSPNIFADVMDDEDAFFDKALEGFALFA 330
 QY 301 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 360
 Db 331 FNOGEVCTCPSRALVQESYERFEMERAIKRVESIRSGNPIDSVTOMGACVSHGQLETTLN 390
 QY 361 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 420
 Db 391 YIDIGKKEGADVLTGRRKLLGELKDGYYLEPTILFGUNNMKVQEEIFGPVLAVITFK 450

QY 421 TMEAEALANDTOYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 |||
 Db 451 TMEAEALANDTOYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 510
 |||
 QY 481 GIGRETHKMMLEHYQOOTKCLLVSYSDKPLGLF 512
 |||
 Db 511 GIGRETHKMMLEHYQOOTKCLLVSYSDKPLGLF 542
 |||

RESULT 4

AF0977
 lactaldehyde dehydrogenase (EC 1.2.1.22) - Salmonella enterica subsp. enterica serovar
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AF0977
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Waig, J.; Churcher, J.; T. Connor, P.; Grodin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fairley, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
 A:Reference number: AB0502; PMID:11677608
 C:Accession: AF0977
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 - PAR
 A:Cross-references: GB AL514982; P1DN:CA007446.1; P1D:q1250445; GSE08 GN00176
 C:Genetics: STY4116
 A:Gene: STY4116
 C:Superfamily: aldehyde dehydrogenase (NAD+) aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 96.9%; Score 2589; DB 2; Length 512;
 Best Local Similarity 95.7%; Pred. No 97-192;
 Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGEYGFLLKARYDNF IGGEWAPADGEYQNLTPVTGQLCEVASSG 60
 |||
 Db 1 MTNPPSTKILQPSXGYPLKARYDNF IGGEWAPADGEYQNLTPVTGQLCEVASSG 60
 |||

QY 61 KRDI LALDAAHKKVKKWAHTSVQKKAALFLKIAURMEQNLLELATAETWNGKPIRETS 120
 |||
 Db 61 KRDI LALDAAHKKVKKWAHTSVQKKAALFLKIAURMEQNLLELATAETWNGKPIRETS 120
 |||

QY 121 AADVPLAIDHFRYFASCIRAOEGGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 |||
 Db 121 AADVPLAIDHFRYFASCIRAOEGGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 |||

QY 181 MAPALAAAGNVVVKPARLTPILSVLLIMFTVGLIPPGVNVVNGAGVIGEYLAISKPIA 240
 |||
 Db 181 MAPALAAAGNVVVKPARLTPILSVLLIMFTVGLIPPGVNVVNGAGVIGEYLAISKPIA 240
 |||

QY 241 KVAFTGSTVEVGQIMQYATONITIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFA 300
 |||
 Db 241 KVAFTGSTVEVGQIMQYATONITIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFA 300
 |||

QY 301 FNGQEVCTTPSPALVQESTYFPFMPFATPPRVESIPSGNPLDSVTOMGAQVSHGQLETILN 360
 |||
 Db 301 FNGQEVCTTPSPALVQESTYFPFMPFATPPRVESIPSGNPLDSVTOMGAQVSHGQLETILN 360
 |||

QY 361 YTDLCKKCAVAVLITGKPKILLESLEKGYLYLEPTILEGJNNMKVFTDEEIPVLAVTTFK 420
 |||
 Db 361 YTDLCKKCAVAVLITGKPKILLESLEKGYLYLEPTILEGJNNMKVFTDEEIPVLAVTTFK 420
 |||

QY 421 TMEAEALANDTOYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 |||
 Db 421 TMEAEALANDTOYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 |||

QY 481 GIGRETHKMMLEHYQOOTKCLLVSYSDKPLGLF 512
 |||
 Db 481 GIGRETHKMMLEHYQOOTKCLLVSYSDKPLGLF 512
 |||

RESULT 5

B87693
 aldehyde dehydrogenase B [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20 Apr 2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: B87693

P.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.F.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Barry, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; H. H.; K. n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc Natl. Acad Sci U S A 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; PMID:21173698; PMID:11259647

A:Accession: B87693

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 - STD

A:Cross-references: GB AE005673; NID:q13425422; P1DN:AAK25542.1; GSPOR:GN00148
 C:Genetics:

A:Gene: CC3580

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 71.8%; Score 1918; DB 2; Length 506;

Best Local Similarity 71.7%; Pred. No. 5.5e-140;

Matches 352; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 22 KARYDNF IGGEWAPADGEYQNLTPVTGQLCEVASSGKKDLALDAAHKKVKKWAHT 81
 |||
 Db 16 KARYDNF IGGEWAPADGEYQNLTPVTGQLCEVASSGKKDLALDAAHKKVKKWAHT 75
 |||

QY 82 SVQGPAAALPKIAUPMEQNLLELATAETWNGKPIRETSAAKVPALTIHFYFASCIKRAQ 141
 |||
 Db 76 SAADRSRILLRTADRMEENLAALATAETWNGKPIRETSAAKVPALTIHFYFASCIKRAQ 135
 |||

QY 142 ECGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWKADALAAAGNVVVKPARLTPIL 201
 |||
 Db 136 ECGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWKADALAAAGNVVVKPARLTPIL 195
 |||

QY 202 SVQLMEVLVGLLPPGVNVVNGAGVIGEYLAISKPIKAVFPISTEVQJIMVATON 261
 |||
 Db 196 SIWVAEMIGDLLPAGVLNIVNGFLEAGKPLASSPKIAFTIGTSTGRLLIMQVAAON 255
 |||

QY 252 IIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFAFNGQEVCTTPSPALVQESTY 321
 |||
 Db 256 IIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFAFNGQEVCTTPSPALVQESTY 315
 |||

QY 322 KMEPALIPKVESIPSGNPLDSVTOMGAQVSHGQLETILNIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFA 381
 |||
 Db 316 KMEPALIPKVESIPSGNPLDSVTOMGAQVSHGQLETILNIPVTLEIGGKSPNIVFADVMDEDAFFDKALEGALFA 375
 |||

QY 382 EGEKIDYVYDPTTIPQJNNMKVFTDEEIPVLAVTTFKTMEPAIPLANIYVGLGAGVW 441
 |||
 Db 376 PQQLADGYVEPTVFEHGNKMRIFQEEIFGVLAVTTFKTEELAEALANDTAFGLGAGVW 435
 |||

QY 442 SPNGNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQSGTGPTTHMMIPLYVQTKCL 501
 |||
 Db 436 SKDANRCYRFGRGTEAGRWTCYHAYPAHAFAFGYKQSGTGPTTHMMIPLYVQTKCL 495
 |||

QY 502 VSYSDKPLGLF 512
 |||
 Db 496 VSYSPKALGFF 506
 |||

RESULT 6

A42597

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) [validated] Alkaligenes eutrophus

N:Alternate names: acetaldehyde dehydrogenase 11

C:Species: Alkaligenes eutrophus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002

C:Accession: A42597

R:Priefert, H.; Jendrossek, D.; Schmidt, B.; Steinbuechel, A.

J. Bacteriol. 174, 899-907, 1992

A:Title: Identification and molecular characterization of the gene coding for acetaldehyde
 A:Reference number: A42597; MUID:92121129; PMID:1732222
 A:Accession: A42597
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <PRI>
 A:Cross-references: GB:M74001, NID:g141885; P10N:AAA21943.1; PID:g141886
 A:Experimental source: strain TF93
 A:Note: sequence extracted from NCBI backbone (NCBIP:77618)
 C:Genetics:
 A:Gene: acd
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: NAD, oxidoreductase
 F:57-321/Domain: aldehyde dehydrogenase homology <ALDD>
 F:262-301/Active site: Glu, Cys #status predicted
 F:456/Binding site: NAD (Cys) #status predicted

Query Match 71.5%; Score 1910; DB 1; Length 506;
 Best Local Similarity 71.5%; Pred. No. 2.3e-139;
 Matches 351; Conservative 67; Mismatches 73; Indels 0; Gaps 0;

QY 22 KARYDNIGGWNAPADGEYQNLTPVTGOLICEVASSCKRDLALDAHKKVKDKWAHT 81
 DB 16 KQYENYIGGAWPPAGGEYFESTTPTKPTPRSRQQQDVDAALDAHAAKAAWART 75
 QY 82 SVODRAATLPTADRMENLELATAFTWDCNKPIRETSAAADVPLAIDHFRYFASCIRAQ 141
 DB 76 STTERANILNRIADRIEANKLLVAESIDNCKPVRETTAADIPLAVDHFYFACIRAQ 135
 QY 142 EGGISEVDSETVAYHFEHPHGVVGGQIIPWNPFLIMASWKMAPALAGNCVLLKPARLTP 201
 DB 136 EGGISEIDADTIAHFEHPHGVVGGQIIPWNPFLIMATWKLAPALAGNCVLLKPAEQTPA 195
 QY 202 SVLLMEIVGDLPPGVVNVVNGAGVIGEYLAISKRIAKVAFSTGVGGQIMQYATON 261
 DB 196 SILVMEIVGDLPPGVVNVVNGAGVIGEYLAISKRIAKVAFSTGVGGQIMQYATON 255
 QY 262 IIPVTELGKSPNIPVADVMEDEDAFFDKALEGFALFAGNCGEYCTCPSRALVGEISYE 321
 DB 256 LIPVTELGKSPNIPVADVMEDEDAFFDKALEGFALFAGNCGEYCTCPSRALVGEISYE 315
 QY 322 REMERARIRVESIRSNPLDSVTOMGAOVSHGQLETLINYYIDIGKKGADVLTCGRKLL 381
 DB 316 RFMERALKRVAALRQGHPLDTGTMIQAQASAFQLEKIIISYIDLGKKGACQCTGERNVL 375
 QY 382 EGELKDGYYLEPTILFQNNMRVFOEIEFGPVLAIVTTKTMEEALELANDTQYGLGAGW 441
 DB 376 DGDLAGGYVVKPTVFAGHKNMRFIOEIEFGPVVSVTTFKDEEALAIANDTLYGLGAGW 435
 QY 442 SRNCLAYKMGRIQAGRWNTNCHYAPAHAAFGGKYGSGIGRETHKMMLEHYQOTKCLL 501
 DB 436 TRGARAFMRGRIQAGRWNTNCHYAPAHAAFGGKYGSGIGRENHMRMLDHYQOTKCLL 495
 QY 502 VSYSDKPLCLF 512
 DB 496 VSYSPNALGPF 506

RESULT 7
 AG2849
 aldehyde dehydrogenase aldA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence,revision 11-Jan-2002 #text,change 01-Feb-2002
 C:Accession: AG2849
 R:Wood, D.W.; Sorubal, J.C.; Kaul, R.; Marks, D.; Chou, L.; Wood, G.E.; Chen, Y.; Woo, I.;
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Woo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AG2577; PMID:11743193
 A:Accession: AG2849

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-505 <KUR>
 A:Cross-references: GB:AF008688; P10N:AA14213.1; PID:g17740695; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: aldA
 A:Map position: circular chromosome
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.9%; Score 1895; DB 2; Length 505;
 Best Local Similarity 69.0%; Pred. No. 3.2e-138;
 Matches 348; Conservative 68; Mismatches 84; Indels 4; Gaps 1;

QY 9 QIKPEGYGPEPLKAKYDINFGGEWVAPALGIEHYQNLTPVTDOLICEVASSCKRDLAL 68
 DB 6 QKAGGAPKPKK--YNYDGKWKVPSGMYMNI SPVIGRKICFVPSNASHDEEAL 61
 QY 69 IAAHKVKDKWAHISVGVRAAATLPTADRMENLELATAFTWDCNKPIRETSAAADVPLA 128
 DB 62 DAAHKAKEKWKTSITERSNILLKIAQRLEDNLLIARAETWDCNKPIRETTNADIP 121
 QY 129 DHFRYFASCIAPAGEGGISEVDSETVAYHFEHPHGVVGGQIIPWNPFLIMASWKMAPALAG 188
 DB 122 DHFRYFAGCIRAOEGTIGEDNUTVAYHFEHPHGVVGGQIIPWNPFLIMAAWKLAPALAG 181
 QY 189 NVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYLAISKRIAKVAFST 248
 DB 182 NVVLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYLAISKRIAKVAFST 241
 QY 249 EVGGQIMQYATONIPVTELGKSPNIPVADVMEDEDAFFDKALEGFALFAGNCGEYCT 408
 DB 242 SVGKEIMRYAADVNTNISLELGKSPNIPVADVMEDEDAFFDKALEGFALFAGNCGEYCT 401
 QY 309 CPSKALVGEISYEFEWEPKALRKVESIRSNPLDSVTOMGAOVSHGQLETLINYYIDIGKKE 468
 DB 302 CPSPALVHESYHREPEKATKPVDAISQDDIPNSTMIQAQASQPFQKIMSYTFPKKE 461
 QY 369 GADVLIGGRKKLEGLKIKYILEPTILFQNNMRVFOEIEFGPVLAIVTTKTMEEALEL 428
 DB 362 GANVLTCGRKLTGDLKDGYYIOPVPSKNMRFIOEIEFGPVVSVTTKTEVEALE 421
 QY 429 ANDTQYGLGAGVWRNCLAYKMGRIQAGRWNTNCHYAPAHAAFGGKYGSGIGRETHK 488
 DB 422 ANDTVYGLGAGVWRNCLAYKMGRIQAGRWNTNCHYAPAHAAFGGKYGSGIGRETHK 481
 QY 489 MMLHYQOTKCLLVSYSKPLCLF 512
 DB 482 MMLDHYQOTKCLLVSYSKPKVGF 505

RESULT 8
 E97626
 chloroacetaldehyde dehydrogenase (AF029733) [imported] - Agrobacterium tumefaciens (S
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence,revision 30-Sep-2001 #text,change 11-Jan-2002
 C:Accession: E97626
 R:Goodner, B.; Binkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Gurolo, R.; Goldm
 A.; Liu, P.; Woliam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
 Science 294, 2323-2328, 2001
 A:Title: Genome sequence of the plant pathogen and biotechnological agent Agrobacterium
 A:Reference number: A97559; PMID:11743194
 A:Accession: E97626
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-540 <KUR>
 A:Cross-references: GB:AF007869; P10N:AAK87966.1; PID:g15151474; GSPDB:GN00169
 C:Genetics:
 A:Gene: AKR_C_4041
 A:Map position: circular chromosome
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.9%; Score 1895; DB 2; Length 540;

Best Local Similarity 69.0%, Pred. No. 1,6e-138;
Matches 348, Conservative 68, Mismatches 84, Indels 4, Gaps 1;

QY 9 QIKRGYGFPLKIKAPYINFTGGPWWAPAGTGYQQNIPTVIRQLDTEVASSCKKROLDAL 68
Db 41 QOKAGEAPFKL---YGNVIGGKWKVEPKSGRYMDNI SPVTGKIKICEVPSDASDFPAL 96
QY 69 DAHKVKKWAHTSVQDPAATLTK (ADPMEONIPLIATARTWONGKPIRETSAADVPALAI 128
Db 97 DAHKAREKWKTSITERSNILLPTAOTFEDNLDLIPAEATWONGKPLPETTNADIPLI 156
QY 129 DHERYFASCTRAQEGGISEVDSEVAYHFEPLGVVGGQITPWNFPLIMASWKMALPAAAG 188
Db 157 DHERYFAGCIRAQEGGISEVDSEVAYHFEPLGVVGGQITPWNFPLIMAAWKLAPALAA 216
QY 189 NCVLKPARLPLSVLLIMETVGLDPLPPGVVNVVNGAGGVIGEVYLAATSKRIAKVARTGST 248
Db 217 NCVLKPAEQTPASILVIMELIEDLPTGVNLVINGTGLDAGPLAGSRIAKIATGST 276
QY 249 EVGGQIMQATONTIPVTLELGGKSNITVFADYMDERDAFFUKALGGFALEAFNQGEVCT 308
Db 277 SVGKEIMEYAAICNVNTISLEIGSKSPNLEFALVNNELTAPLLKALGGFALEAFALNQGEVCT 348
QY 309 CPSRAI VCPSEIYFFHMPAPAPPVESIPSGNPILSVTQMGAGVSHQIETIIINVDGKKE 368
Db 337 CPSRAI VCHESIYFFHMPKAIKPVQALISQDEPLNPSMTMGASQDEPLKIMSYLEIGKKE 408
QY 369 GADVTIGCPPKLIGSELKGGYILEPTILEGNNNNPVFOEIPGPVLAVTTTEKTMEEALEL 428
Db 397 GAKVLTGGDKLTIGDLKGGYIQPTVFGNNKKIKFOEDIPGPVVSVTTEKKEVDALEI 456
QY 429 ANDTOYGGIAGVSWRNQNIAYKMGRTGTQAGVWNTCYHAYPAHAAPGGYKQSGIRETHK 488
Db 457 ANDTVIGLGGVSWRNDTNIAYKMGRTGTQAGVWNTCYHAYPAGAAFGYKQSGIRETHK 516
QY 489 MWLEHYQKIKCLLVSYSDKPLGLF 512
Db 517 MWLDHYQKIKCLLVSYSPNKVGF 540

RESULT 9
A83735
aldehyde dehydrogenase aldA [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A83735
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83735
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <STO>
A:Cross-references: GB:AP001509; GR:BA000004; NID:g10173176; PIRN:PAR04400.1; GSPDR:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.3%, Score 1878; DB 2; Length 506;
Best Local Similarity 68.9%, Pred. No. 6.7e-137;
Matches 346; Conservative 67; Mismatches 89; Indels 0; Gaps 0;

QY 11 KPGEYGFPLKIKARYDNFTGTGWWAPAGTGYQQNIPTVIRQLDTEVASSCKKROLDALDA 70
Db 5 RPNEGAVSVFKKRYDNTVNGEWTTPVKGYFENVTPVTGEVCEVARTDAEDLALDA 64
QY 71 AHKVKKWAHTSVQDRAILFKIADRMQNLLELATAETWONGKPIRETSAADVPALAI 130
Db 65 AHAKFQWKTSPAPRANTLNKTADPMEENLEKLAETWONGKPIRETSAADVPALAI 124
QY 131 FRYFASCTRAQEGGISEVDSEVAYHFEPLGVVGGQITPWNFPLIMASWKMALPAAAGNC 190

Db 125 FRYFAGALRAVGTLSDINDVYHFEPLGVVGGQITPWNFPLIMASWKMALPAAAGNC 184
QY 191 VVIFAPALPTLSVLI LMFIVGDL PFGVWNVVNGAGVIGYQIATSPFIAYVAFTGSTEV 250
Db 185 VILKPAEQTPASIFVLELIEDLTPKGVNINLVSPGVGAGKPLASSPSAKVAFGETTIT 244
QY 251 QQQIMQATONTIPVTLELGGKSNITVFADYMDERDAFFUKALGGFALEAFNQGEVCTCP 310
Db 245 GRLIMQVASENLIPVTLELGGKSNIPFDDVMDKDDAFUNKAIEGVLEFALNQGEVCTCP 304
QY 311 SPALVQESIVYERFMEPAIPKVESIPSNPLSVLMAGVASHQIETIIINVDGKKE 370
Db 305 SKALLIQESIVDTFMERALAKVQALKQGNPLDPNTMIGAAQASQDEPLKIMSYLEIGKKE 364
QY 371 DVLTCGRKKLLELGGKSYLEPTILEGNNNNPVFOEIPGPVLAVTTTEKTMEEALELAN 430
Db 365 EVLAGGGRNHLFGELANGYVSPVTFKGTNDKMRVFOEIPGPVVSVTTEKDAEEALAIAN 424
QY 441 DLQVGLGAGVSWRNQNIAYKMGRTGTQAGVWNTCYHAYPAHAAPGGYKQSGIRETHKMM 490
Db 445 GLLVGLGAGVWTPKMKKAYPPKPELQAGVWNTCYHAYPAHAAPGGYKQSGIRETHLMM 484
QY 491 LERHYQKIKCLLVSYSDKPLGLF 512
Db 445 LERHYQKIKCLLVSYSPNKVGF 506

RESULT 10
H75589
aldehyde dehydrogenase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75589
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIRN:AAF12436.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0348
A:Map position: 2
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 69.2%, Score 1849.5; DB 2; Length 515;
Best Local Similarity 69.1%, Pred. No. 1.1e-134;
Matches 346; Conservative 59; Mismatches 95; Indels 1; Gaps 1;

QY 12 KGEYGFPLKIKARYDNFTGTGWWAPAGTGYQQNIPTVIRQLDTEVASSCKKROLDALDA 71
Db 16 KGFPGSVVTFKKRYDNTFVGWVFPVKGYFENVTPVTGEVCEVARTDAEDLALDA 75
QY 72 HKVKRWKWAHTSVQDRAILFKIADRMQNLLELATAETWONGKPIRETSAADVPALAI 131
Db 76 HPAAPAWKSPISVIFKSNILKIALPMEQNLLELATAETWONGKPIRETSAADVPALAI 135
QY 132 PYFASCTRAQEGGISEVDSEVAYHFEPLGVVGGQITPWNFPLIMASWKMALPAAAGNCV 191
Db 136 PYFAGCTRAQEGGISEVDSEVAYHFEPLGVVGGQITPWNFPLIMASWKMALPAAAGNCV 195
QY 192 VLKPARLPLSVLLIMETVGLDPLPPGVVNVVNGAGGVIGEVYLAATSKRIAKVARTGSTEV 251
Db 196 VLKPAEQTPASIFVLELIEDLTPKGVNINLVSPGVGAGKPLASSPSAKVAFGETTIT 255
QY 252 QQQIMQATONTIPVTLELGGKSNITVFADYMDERDAFFUKALGGFALEAFNQGEVCTCP 311
Db 252 QQQIMQATONTIPVTLELGGKSNITVFADYMDERDAFFUKALGGFALEAFNQGEVCTCP 311

Db 256 RLIMGYAADNLIPVTELGKSPNIFDDVMEDDAFLDKAVEGVMMFALNGEVCTCP 315
 QY 312 RALVQESIYERFPRERARRVESTIRSNPLDSVTGMAQVSHQOETILNYIDIGKKEGAD 371
 Db 316 RALVQESIYERFPRERARRVESTIRSNPLDSVTGMAQVSHQOETILNYIDIGKKEGAD 375
 QY 372 VLTGGRKLLLEGKIDGYYLEPIFGQNNMRVFOEIEFGPVLAVTTFKTMEEALELAND 431
 Db 376 VLTGGRKLLLEGKIDGYYLEPIFGQNNMRVFOEIEFGPVLAVTTFKTMEEALELAND 434
 QY 432 TOYGIAGVMSRGNLAYKMGRIQAGRVNTNCYHAYPAHAAGGYKQSGIGRTHKMWL 491
 Db 435 TLYGLAGTWTROISPAYRMGRIGQAGRVNTNCYHAYPAHAAGGYKQSGIGRTHKMWL 494
 QY 492 EHYQOTKLLVSYSDKPLGLP 512
 Db 495 DHYQOTKLLVSYSDKPLGLP 515

RESULT 11

AE3470
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) [imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-May-2002
 C:Accession: AE3470
 R:DelVecchio, V.G.; Kaparral, V.; Pedkar, P.T.; Patra, G.; Mujar, C.; Los, T.; Ivanova, A.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen brucella melitensis
 A:Reference number: ANR352; PMID:11756688
 A:Accession: AE3470
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-505 <KUR>
 A:Cross-references: GB:AE008417; PIRN:AA152528 1; PIR:317984776; QSPR:Q000140
 A:Experimental source: strain 16M
 C:Genetics:
 A:Map position: 1
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 69.2%; Score 1848.5; DB 2; Length 505;
 Best Local Similarity 67.4%; Pred. No. 1,36-134;
 Matches 341; Conservative 65; Mismatches 93; Indels 7; Gaps 1;

QY 7 SAOTKPGEGEPLKARYDNFEGGEWVAPADGEYQNLTPVTGQLGCEVASSKRRDIDLAA 66
 Db 7 SRVAKPA-----PAKRYCNFEGKWKVPPSPRPENTSPVNGGLVIFVAFSTAAVFA 59
 QY 67 ALDAHKVKIKWAHTSVQDPAALPKIADPMQNLLELLATAETWONGKPIRETSAAVPL 126
 Db 60 ALDAHAARELWGRISVAERAILILNIRADRIENLPALAAETWONGKPIRETTMADLPL 119
 QY 127 AIDHRYFASCIQAQEGGISEVDSEVAVHFEHPGVVGVQIIPWNPFLLMASWMAPALA 186
 Db 120 AVDHRYFACVIRQAQGGRISEIDHDFVAVHFEHPGVVGVQIIPWNPFLLMATWKLAPALA 179
 QY 187 AGNCVVLKPARLPIPLWLLMELIVSDIIPPGVNVVNVNGAGVIGVGLATSKRIKAVFTG 246
 Db 180 AGNCVVLKPAEQTPASILMELIADIPPGVNVVNVNGGLENGKPLASSPRIAKIATG 239
 QY 247 STEVGQIIMOYATQNIIPVTLELGKSPNIVAIAMVDEDAFPAKALGFALFAPNGGV 306
 Db 240 EITTGRIIMQYASQNIIPVTLELGKSPNIVAIAMVDEDAFPAKALGFALFAPNGGV 299
 QY 307 CTCPSRALVQESIVEREMERARRVESIRSGNPLDSVTGMAQVSHQOETILNYIDIGK 366
 Db 300 CTCPSRALVQESIVEREMERARRVESIRSGNPLDSVTGMAQVSHQOETILNYIDIGK 359
 QY 367 KEGADVLTCGRKRLLEGKIDGYYLEPIFGQNNMRVFOEIEFGPVLAVTTFKTMEEAL 426
 Db 360 QEGAELVAGNERMNLGGLAGGYVKKPTVFKHNMKRIQEIEFGPVVSVATPKDDAAL 419

QY 427 ELANDTQYGLGAGVMSRGNLAYKMGRIQAGRVNTNCYHAYPAHAAGGYKQSGIGRTH 486
 Db 420 ELANDTQYGLGAGVMSRGNLAYKMGRIQAGRVNTNCYHAYPAHAAGGYKQSGIGRTH 479
 QY 487 HKMMLHYQOTKLLVSYSDKPLGLP 512
 Db 480 HKMMLHYQOTKLLVSYSDKPLGLP 505

RESULT 12

F83142
 probable aldehyde dehydrogenase PA4022 [imported] Pseudomonas aeruginosa (strain PA
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 41-Dec-2000
 C:Accession: F83142
 R:Stover, C.K.; Pham, X.Q.; Brown, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathin, R.J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: AB2950; PMID:2047337; PMID:10984043
 A:Accession: F83142
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <STO>
 A:Cross-references: GB:AE004819; GB:AE004091; NID:q9950214; PIRN:AAGU7409.1; QSPR:Q000140
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4022
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 67.4%; Score 1802; DB 2; Length 506;
 Best Local Similarity 67.1%; Pred. No. 4,96-141;
 Matches 336; Conservative 59; Mismatches 106; Indels 0; Gaps 0;

QY 12 PGEYGFPLKARYDNFEGGEWVAPADGEYQNLTPVTGQLGCEVASSKRRDIDLAA 71
 Db 6 PGEYGFPLKARYDNFEGGEWVAPADGEYQNLTPVTGQLGCEVASSKRRDIDLAA 65
 QY 72 HKYDKWAHTSVQDPAALPKIADPMQNLLELLATAETWONGKPIRETSAAVPLADHP 141
 Db 66 HAAPAWKRPSTVQSPNIIKIAPIEQLNFIIVTETWONGKPIRETSAAVPLADHP 126
 QY 132 KTFASCIQAQEGGISEVDSEVAVHFEHPGVVGVQIIPWNPFLLMASWMAPALA 191
 Db 126 KTFASCIQAQEGGISEVDSEVAVHFEHPGVVGVQIIPWNPFLLMASWMAPALA 185
 QY 192 VLFPAVLTLSVLLMELIVSDIIPPGVNVVNVNGAGVIGVGLATSKRIKAVFTG 251
 Db 186 VLFPAVLTLSVLLMELIVSDIIPPGVNVVNVNGAGVIGVGLATSKRIKAVFTG 245
 QY 252 QIIMOVATQNIIPVTLELGKSPNIVAIAMVDEDAFPAKALGFALFAPNGGV 411
 Db 246 QIIMOVATQNIIPVTLELGKSPNIVAIAMVDEDAFPAKALGFALFAPNGGV 405
 QY 412 RALVQESIYERFPRERARRVESTIRSNPLDSVTGMAQVSHQOETILNYIDIGKKEGAD 471
 Db 406 RALVQESIYERFPRERARRVESTIRSNPLDSVTGMAQVSHQOETILNYIDIGKKEGAD 465
 QY 472 VLTGGRKLLLEGKIDGYYLEPIFGQNNMRVFOEIEFGPVLAVTTFKTMEEALELAND 441
 Db 466 VLTGGRKLLLEGKIDGYYLEPIFGQNNMRVFOEIEFGPVLAVTTFKTMEEALELAND 426
 QY 492 EHYQOTKLLVSYSDKPLGLP 512
 Db 486 EHYQOTKLLVSYSDKPLGLP 506

RESULT 13

F70827

Probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) P:0458 [similarity] - Mycobacterium

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 04-Jun-2002

C:Accession: F70827

R:Cole, S.T.; Brochu, P.; Parkhill, J.; Garnier, T.; Churrier, C.; Harris, D.; Gordon, S.

J. Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

C:Accession: F70827

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-507 <COL>

A:Cross-references: AB:AL027633, AB:AL123456, NID:9356129; PIR:AA17413.1, PIR:Q20954

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV0458

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: oxidoreductase

F:58-322/Domain 3:3-hyd- dehydrogenase homology: AL5-

RESULT 14

H83396

Probable aldehyde dehydrogenase PA1984 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 16-Sep-2000 #seq-prot_revision 16-Sep-2000 #text_change 31 Dec 2000

C:Accession: H83396

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickney, M.J.; R

adnato, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbia, K., L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82350; MUID:2043747; PIR:J0994043

A:Accession: H83396

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <ST>

A:Cross-references: GR:AF004625, GR:AE004091, NID:99947983, PIR:AA655372.1, GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1984

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

A: Experimental source: serogroup O1; strain N16961; biotype El Tor

C: Genetics:

A: Gene: VC1819

A: Map position: 1

C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```
Query Match          67 0%; Score 1791; PR 2; Length 506;
Best Local Similarity 64.9%; Pred No. 3.4e-130;
Matches 326; Conservative 74; Mismatches 102; Indels 0; Gaps 0;

QY 11 KPGRYGFPFLKARKYDNFICGEWAPADGEYQYQNLTPVTCQLICEVASSGKKRDIDLALDA 70
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 QPGSDNAVITFKSHYDNFICGQWVKVSGEYFQNFSEVNGQYVQGVAKSTQADIDLALDA 64

QY 71 AHKYKDWAKTWSQDRAAILFKIADRMFQNLLELLATAETWDCNCKPIRETSAADVPLAIDH 130
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 AHQVREAWAKTSTVTERSNLLKIDRIEANIQLVAECWENGKPVRETAAADLPVVVDH 124

QY 131 FRYFASCIQAEGGISEVDSQETVAYHFHEPLGVVGQIIPWNPFLMASMKMAPALAAAGNC 190
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 FRYFAGCIQAEGSAAELDSHTASYHFPPEPIGVVGQIIPWNPFLMAAKLAPALAAAGCC 184

QY 191 VVLKPARLTPLSVLLMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIAKVAFTGSTEV 250
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 VVLKPAEQTPSTILVLEKTIADLIPAGVLNVVNGFSGEAGQALATSQRIAKLAFTGSTQV 244

QY 251 GQQLMQVATQNIIPVILEAGKSPNIVFAVLMUEUAFDQKALEGFALFAFNGGVCVCP 310
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 GQHILKCAOSLIPSTVEIGCKSPNIYFPDIFDHEDTYLEKCIETLLGFFNQGVCVCP 304

QY 311 SRALVQESIYERFMERAIRVESIRSGNPLDSVTOMGAQVSHGQLETILNYIDICKKEGA 370
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 SRVLVHESIYDRFVAKVAERAKGIKQGNPLDTATQVGAQASQFQPKTILSYIDIGRQCA 364

QY 371 DVLTCGRKKLLEGELKNGYVLEPTILFGONNMHVFOEITFGPVIAVTTFKTMEEALELAN 430
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
365 KVLGGGVAKQEGELGGYTIQPTLLOGHNMKRVFOEITFGPVIAITSPKDEAEALAN 424

QY 431 DTQYGLGAGVWSRNCNIAYKMGFGIQAGPVWNTNCHAYIPAHAAFGGYKSGIGRETHKMM 490
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 DTEYGLGAGIWTWQUNLAYRMGNLIQAGRIWNCYHAYIPAHAAFGGYKSGIGRETHKMM 484

QY 491 LEHYQQTCKLLVSYSDKPIGLF 512
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 LNHVQNTKNLLISYDVNPIGLFF 506
```

Search completed: June 24, 2003, 10:18:33
Job time : 21.5932 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:55 ; Search time 10.3122 seconds

(without alignments)
2059.300 Million cell updates/sec

Title: US-09-830-751-8

Perfect score: 2673

Sequence: 1 MNNPSSAQIKPGYGFPLK

HYQQTCKLLVSYSEKPLGLF 512

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 4147628 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2668	99.8	512	1 ALDB_ECOLI	P37685 escherichia
2	1910	71.5	506	1 DH2A_ALCFU	P46368 alcaigenes
3	1849	69.2	515	1 DH2A_DEIRA	Q97999 deinoxococcus
4	1810	67.7	506	1 DH2A_RHORI	Q92411 rhodospirillum rubrum
5	1798	67.3	507	1 DH2A_MYCTU	O53743 mycobacterium tuberculosis
6	1764	66.2	505	1 THCA_PHOEF	P46369 rhodococcus ruber
7	1760	65.8	507	1 DH2A_STRCO	Q97326 streptomyces coelicolor
8	1682	62.4	506	1 DH2A_VLHCH	P23240 vibrio cholerae
9	981	36.7	519	1 DH2A_MOUSE	P47738 mus musculus
10	975	36.5	519	1 DH2A_RAT	P11884 rattus norvegicus
11	971	36.3	517	1 DH2A_HUMAN	P05091 homo sapiens
12	968	36.2	500	1 DH2A_MESAU	P81178 mesocricetus auratus
13	953	35.7	520	1 DH2A_BOVIN	P20900 bos taurus
14	952	35.6	500	1 DH2A_HORSE	P12762 equus caballus
15	940.5	35.2	519	1 DH2A_YEAST	P46367 saccharomyces cerevisiae
16	936	35.0	517	1 DH2A_HUMAN	P30837 homo sapiens
17	931.5	34.8	499	1 DH2A_HUMAN	Q64768 homo sapiens
18	925.5	34.8	497	1 DH2A_HUMAN	P08157 emericella
19	925.5	34.8	500	1 DH2A_HUMAN	Q74187 agaricus bisporus
20	925.5	34.7	499	1 DH2A_HUMAN	Q63639 rattus norvegicus
21	925.5	34.6	499	1 DH2A_HUMAN	Q63148 mus musculus
22	922	34.5	519	1 DH2A_MOUSE	P40047 saccharomyces cerevisiae
23	921.5	34.5	499	1 DH2A_HUMAN	Q63344 gallus gallus
24	910	34.0	490	1 DH2A_HUMAN	P47805 homo sapiens
25	905	33.9	512	1 DH2A_HUMAN	P41751 aspergillus fumigatus
26	903.5	33.8	497	1 DH2A_HUMAN	P27453 gallus gallus
27	895	33.5	509	1 DH2A_HUMAN	Q27640 enchytraeus albidus
28	881	33.0	497	1 DH2A_HUMAN	P44115 saccharomyces cerevisiae
29	880.5	32.9	499	1 DH2A_HUMAN	P00352 homo sapiens
30	873.5	32.7	500	1 DH2A_HUMAN	P15437 equus caballus
31	870.5	32.6	500	1 DH2A_HUMAN	P42041 alternaria alternata
32	870	32.5	495	1 DH2A_HUMAN	Q64174 oryza sativa
33	868	32.5	505	1 DH2A_HUMAN	

ALIGNMENTS

RESULT 1

ALDB_ECOLI

ID ALDB_ECOLI STANDARD; PRT; 512 AA.

AC P37685; P78118;

DT 01-OCT-1994 (Rel. 30, Created)

DI 01 NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Aldehyde dehydrogenase B (EC 1.2.1.22) (Lactaldehyde dehydrogenase).

GN ALDB OR B3588.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RP [1]

RN SEQUENCE FROM N.A.

PC STRAIN-K12 / MG1655;

PX MPD.INF=94316500; PubMed=8041620;

RA Sofia H.J., Burland V., Daniels B.L., Plunkett G. III, Blattner F.R.;

RT Analysis of the Escherichia coli genome. V. DNA sequence of the

region from 76.0 to 81.5 minutes.;

PL Nucleic Acids Res 22:2576-2586(1994).

RN [2]

RP SEQUENCE FROM N.A.

PC STRAIN-K12 / MG1655;

PX MEDLINE=95286498; PubMed=7768815;

RA Xu J., Johnson R.C.;

RT "aldB, an Rpos dependent gene in Escherichia coli encoding an

aldehyde dehydrogenase that is repressed by Fis and activated by

Crp.;"

RL J. Bacteriol. 177:3166-3175(1995).

CC : CATALYTIC ACTIVITY: (S)-Lactaldehyde + NAD(+) + H(2)O = (S)-

lactate + NADH.

CC -! SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

CC This SWISS-Prot entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.sit.ch/announce/> or send an email to license@isb.sit.ch).

EMBL: 000039, AAR18565.1; ALT_INIT.

EMBL: AF030436, AAC76612.1; ALT_INIT.

EMBL: L40742, AAC36939.1; -

HSSE; F05091; ICM3.

CC EMBL: E01292, ALDB

CC Information: P46368; Aldehyde dehydrogenase.

CC Pfam: PF001071; aldedh; 1.

CC ProSITE: PS00687; ALDEHYDE_DEHYDR_GYS; 1.

CC ProSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.

CC Oxidoreductase, NAD, Complete proteome.

FT ACT_SITE 268 268 BY SIMILARITY.

FT ACT_SITE 307 307 BY SIMILARITY.

SEQUENCE 512 AA: 56306 MW; 6573C34340DF68CD CRC64;

```

Query Match      99.8%   Score 2668   DB 1: Length 512;
Best Local Similarity 99.8%   Pred No 1 2c 192;
Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGEGYFPIKIKARYINFGKRWVAPAIKGYUNLTPVTSGLLEEVASSG 60
DB 1 MTNPPSAQIKPGEGYFPIKIKARYINFGKRWVAPADGEYUNLTPVTGQLLCHVASSG 60

QY 61 KPDIDALDAAHKVKIKWAHTSVQDPAALFKLAIHPMFUNLELLATAETWNGKPIRETS 120
DB 61 KPDIDALDAAHKVKIKWAHTSVQDPAALFKLAIHPMFUNLELLATAETWNGKPIRETS 120

QY 121 AADVPLAIDHFRYFASCIKRAOEGISEVDSEIVAYHFHPEPLGVGQIIPWNPPLMASWK 180
DB 121 AADVPLAIDHFRYFASCIKRAOEGISEVDSEIVAYHFHPEPLGVGQIIPWNPPLMASWK 180

QY 181 MAPAAGNCVWLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYLATSKRIA 240
DB 181 MAPAAGNCVWLKPARLTPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYLATSKRIA 240

QY 241 KVAFTGSTEVGOOIMOVATONITPVTLLEGGKSPNIVFADVMDDEDAFDDKALEGFALFA 300
DB 241 KVAFTGSTEVGOOIMOVATONITPVTLLEGGKSPNIVFADVMDDEDAFDDKALEGFALFA 300

QY 301 FNOGEVCTCPSRALVQESIVEREMERAIIPRVESIRSGNPIDSVTOMGAQVSHGOLETILN 360
DB 301 FNOGEVCTCPSRALVQESIVEREMERAIIPRVESIRSGNPIDSVTOMGAQVSHGOLETILN 360

QY 361 YIDIGKKEGAVLTGGRKKLLEGLKIGYYLEPTILFGUNNNRVFUEIFGVPLAVTTFK 420
DB 361 YIDIGKKEGAVLTGGRKKLLEGLKIGYYLEPTILFGUNNNRVFUEIFGVPLAVTTFK 420

QY 421 TMEALELANUTGYLGAGVWSKNGNLAYKMGKGLUAGVWLNCHYHAYPAHAAPFGYKQS 480
DB 421 TMEALELANUTGYLGAGVWSKNGNLAYKMGKGLUAGVWLNCHYHAYPAHAAPFGYKQS 480

QY 481 GIGRETHKMMLEHYQOTKCLLSVSDKPLGLF 512
DB 481 GIGRETHKMMLEHYQOTKCLLSVSDKPLGLF 512

RESULT 2
DH2A2_ALCUO STANDARD; PRI; 506 AA.
AC P46368;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetaldehyde dehydrogenase II (acoh) of Alcaligenes eutrophus.
GN ACOD.
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RX MEDLINE=92121129; PubMed 1732222;
RA Pfeiffer H, Krueger N, Jendrossek D, Schmidt B., Steinbuechel A.;
RT "Identification and molecular characterization of the gene coding for
RL J. Bacteriol. 174:999-1007 (1992).
CC 1-1- FUNCTION: INVOLVED IN THE CATABOLISM OF ACETOIN AND ETHANOL.
CC 1-1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC 1-1- PATHWAY: Ethanol utilization, second step.
CC 1-1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch)

```

```

CC FMRL M74093; AAA21943.1;
DR HSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF06171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NTRINP 246 245 NAD (NUP PART) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 301 301 BY SIMILARITY.
SQ SEQUENCE 506 AA; 54881 MW; A8715R094B126MD CR64;

Query Match      71.5%   Score 1910; DB 1: Length 506;
Best Local Similarity 71.5%   Pred No 8.7e-146;
Matches 351; Conservative 67; Mismatches 74; Indels 0; Gaps 0;

QY 22 KARYDNFTGGEVAPADGEYUNLTPVTGQLLCEVASSCKRIDLALDAAHKVKIKWAHT 81
DB 16 KOOVENYIGCAVPPAGGEYFESTTPTCKPPTVRVRSQQDVAALAAHAAKAAWAKT 75

QY 82 SVQDRAALIFK IADRMQNLLELATAETWNGKPIRETSAADVPLAIDHFRYFASCIKRA 141
DB 76 SITERANILNRIADRIEANKLIAVAESIDNGKIVRETTAAADPLAVIHFRYFAGITRAQ 135

QY 142 EGGISEVDSEIVAYHFHPEPLGVGQIIPWNPPLMASWKMAIPALAGN-VVLIKARETPI 201
DB 136 EGGISELDAIT IAYHFHPEPLGVGQIIPWNPPLMATWKLALALAGN-VVLIKARETPI 195

QY 202 SVLLMEIVGDLPPGVVNVVNGAGVIGEYLATSKRIA KVAFTGSTEVGOOIMOVATON 261
DB 196 SVLLMEIVGDLPPGVVNVVNGAGVIGEYLATSKRIA KVAFTGSTEVGOOIMOVATON 255

QY 262 LIPVTLEGGKSPNIVFADVMDDEDAFDDKALEGFALFAVWLNCHYHAYPAHAAPFGY 421
DB 256 LIPVTLEGGKSPNIVFADVMDDEDAFDDKALEGFALFAVWLNCHYHAYPAHAAPFGY 415

QY 322 PMEPALIPRVESIRSGNPIDSVTOMGAQVSHGOLETILNFGUNNNRVFUEIFGVPLAV 481
DB 316 PMEPALIPRVESIRSGNPIDSVTOMGAQVSHGOLETILNFGUNNNRVFUEIFGVPLAV 475

QY 382 EGGISEVDSEIVAYHFHPEPLGVGQIIPWNPPLMASWKMAIPALAGN-VVLIKARETPI 441
DB 376 EGGISEVDSEIVAYHFHPEPLGVGQIIPWNPPLMASWKMAIPALAGN-VVLIKARETPI 435

QY 442 SRNGLAYKMGKGLUAGVWLNCHYHAYPAHAAPFGYKQSIGRETHKMMLEHYQOTKCLL 501
DB 436 TRDCARAFRMGKGLUAGVWLNCHYHAYPAHAAPFGYKQSIGRETHKMMLEHYQOTKCLL 495

QY 502 VSYSDKPLGLF 512
DB 496 VSYSPNALGFF 506

RESULT 3
DH2A2_DEIRA STANDARD; PRI; 515 AA.
AC Q9Y939; O42502;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase (Bc 1.2.1.4).
GN ALDA OR DRA0448.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/deinococcus group; Deinococci; Deinococcaceae;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1.
PX MEDLINE=20936896; PubMed=19567266;
PA White G., Fison J.A., Heidebolter J.F., Hickey F.K., Petersson J.D.,
PA Dodson R.J., Hatt D.R., Gwyn M.L., Nelson W.C., Richards D.L.,

```


RT dehydrogenase.":
 RL J. Bacteriol. 177:676-687(1995).
 CC !- FUNCTION: DEGRADATES ALL ALDEHYDES POTENTIALLY GENERATED BY N
 CC DEALKYLATION OF THIOCARBAMATES AND MAY ALSO PARTICIPATE IN
 CC ETHANOLAMINE METABOLISM AND FURTHER ASSIMILATION OF DEGRADATION
 CC PRODUCTS BY THIOCARBAMATE-INDUCED CYTOCHROME P-450
 CC !- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC !- INDUCTION: BY EPTC (S-ETHYL DIPHOSPHYLICAPRAMOTHYLALF).
 CC !- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U17129; AAC77472.1; .
 CC HSSP: P05091; 1CW3.
 CC InterPro: IPR002086; Aldehyde_dehydr.
 CC Pfam: PF00171; alde-deh. 1.
 CC PROSITE: PS00670; ALDEHYDE_DEHYDR_CYS; 1.
 CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 CC Oxidoreductase; NAD.
 CC KW INIT_MET 0 0
 CC FT NP_BIND 218 224 NAD (ADP PART) (BY SIMILARITY).
 CC FT ACT_SITE 262 262 BY SIMILARITY.
 CC FT ACT_SITE 301 301 BY SIMILARITY.
 CC SEQUENCE 505 AA; 54863 MW; DBEDEC38B34B37 CRC64;

Query Match 66.2%; Score 1768.5; DR 1; Length 505;
 Best Local Similarity 65.9%; Pred. No. 3,6e-125;
 Matches 331; Conservative 68; Mismatches 102; Indels 1; Gaps 1.
 QY 11 KPGEYGFPLKLRKYDNFICGEWVAPADGYYQNIPTVTGOLLCEVASSCKRPDIALDA 70
 DB 5 KPGTADATMSQYDQWICNGENWAPVKGQYFENFPTVQGNCPHWASIAEDIELDA 64
 QY 71 ARKVKDKWHTSVQDRAAILFKIADRMQNIELIATFWDNGKPIKETSAAIDVPLAIDH 140
 DB 65 AHAAAPANGKTSVAEPAITLNTKATPMEKNLESIALAESWONKPIPETLNAIPLAIDH 124
 QY 131 PRYFASCIKRAEGGKISGVDSHTVAYHPEHPIGVVQGIIPWNPFFILMASWMAFALAAGN 140
 DB 125 FRYFAGATRAQEGSLSEINSDTVAYHPEHPIGVVQGIIPWNPFFILMAVVKLAPALAA 184
 QY 191 VLKPARLTPLSVLLMEIVGDLPLPGVNVVNCAGGVIGEYLATSKRIAKVAFTGSTEV 250
 DB 185 IVLKPAECTPSVILHLGICGLDLPAGVLNVNGFVEAGKPLASSPRILKTAFTGETIT 244
 QY 251 GQIQMQYATQNIIPVTELEGGKSNIVFADVMDDEDAFFDKALEGFPALFAPNOGEVCTCP 310
 DB 245 GRLLIMOVASQNLIPVTELEGGKSNVFPFSDVLASNDYDQKALEGFTMEALNOGEVCTAP 304
 QY 311 SRALVQESIYERPFERAIKRVESIRSNPLDSVTOMCAOVSHCOLETLINYNIDICKEGA 370
 DB 305 SRALLIQEDIDFEFLAMAAIRTKAVRQGGPIITDTIMCAQNSDQLEKILSYIEIGKAGA 364
 QY 371 DVLTGGRKRLIFGLIKMYVTEITLFGQNNMVFQGFIFGSPVLAVTTEKTEMEEALELAN 430
 DB 365 KVIITGGPARKGGPISGYYVQPTVFTFGNKNMPTFQ-EIFGPPVSVTSFKDYDEALIEIAN 423
 QY 431 DTQVGLGNGVSWSPNCLAYKMGFTIQAGFQWNTWYHAYPAHAAPFGYFQSGIIPFTHKMM 440
 DB 424 ETLVSLGAGVMSREGSVAYFAGKULQAGFVMIINTYHQYPAHAAPFGYFQSGIGRENHLM 443
 QY 491 LEHYQQTQKCLLVSYSDKPLGLF 512
 DB 484 LSHYQQTQKNLLVSYAQAQAGFF 505

RESULT 7

DEHAL_STRCO STANDARD; PPT: 507 AA
 ID DEHAL_STRCO
 DT Q9R3Z6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable aldehyde dehydrogenase (EC 1.2.1.3).
 GN S01174 OR S031A.05.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SOURCE: FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2190410, PubMed=2000955,
 RA Bentley S.D., Chater K.F., Gerdes-Tarraga A.-M., Challis G.L.,
 RA Thomson N.P., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke I., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Senger K., Saunders D., Sharp S., Squares P., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrall H.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)."
 EL Nature 417:141-147(2002).
 CC !- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH
 CC !- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL143210; CAB61586.1; .
 CC HSSP: P05091; 1CW3.
 CC InterPro: IPR002086; Aldehyde_dehydr.
 CC Pfam: PF00171; aldehyd; 1.
 CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 CC PROSITE: PS00670; ALDEHYDE_DEHYDR_CYS; 1.
 CC Oxidoreductase; NAD; Complete proteome.
 CC NP_BIND 219 225 NAD (ADP PART) (BY SIMILARITY).
 CC FT ACT_SITE 263 263 BY SIMILARITY.
 CC FT ACT_SITE 302 302 BY SIMILARITY.
 CC SEQUENCE 507 AA; 55673 MW; 0110B15B2D8F96DD CRC64;
 Query Match 65.8%; Score 1760; DR 1; Length 507;
 Best Local Similarity 65.9%; Pred. No. 1,6e-124;
 Matches 330; Conservative 64; Mismatches 107; Indels 0; Gaps 0;
 QY 12 PGEYGFPLKLRKYDNFICGEWVAPADGYYQNIPTVTGOLLCEVASSCKRPDIALDA 71
 DB 7 PGTETALIVSYSPYDHPITGTYVPPAPGQYFENFPTVQGNCPHWASIAEDIELDA 66
 QY 72 HKVVDKDKWHTSVQDRAAILFKIADRMQNIELIATFWDNGKPIKETSAAIDVPLAIDH 131
 DB 67 HEAAPGQPTSVTEPSITLKIADRMQNIELIATFWDNGKPIKETSAAIDVPLAIDH 126
 QY 142 PYFASCIKRAEGGKISGVDSHTVAYHPEHPIGVVQGIIPWNPFFILMASWMAFALAAGN 191
 DB 127 PYFAGATRAQEGSLSEINSDTVAYHPEHPIGVVQGIIPWNPFFILMAVVKLAPALAA 186
 QY 192 VLKPARLTPLSVLLMEIVGDLPLPGVNVVNCAGGVIGEYLATSKRIAKVAFTGSTEV 251
 DB 187 VLKPAECTPSVILHLGICGLDLPAGVLNVNGFVEAGKPLASSPRILKTAFTGETIT 246
 QY 252 GQIQMQYATQNIIPVTELEGGKSNIVFADVMDDEDAFFDKALEGFPALFAPNOGEVCTCP 311
 DB 246 GRLLIMOVASQNLIPVTELEGGKSNVFPFSDVLASNDYDQKALEGFTMEALNOGEVCTAP 304

DB 247 RLIMQYASENIKPVILEGKSNIFEDVWARDDDFROKALEGTFMFALNOGEVCTGCS 306
 QY 312 RALVQESIYERFERAIRRVESIRSGNPLISVTCMAQVSHQULETILNYIUIGKKEAD 371
 DB 307 RALVORGWYAFMEAAVARTELKPGHPLDPTDMICQAQSNQOLEKILSYLDIGROGAK 366
 QY 372 VLTGCRKILGELKDGYYLEPTILFGQNNMRVFOEIEFGPVLAVTTEKTEEALAND 431
 DB 367 VLTGGERLEHD:ELKGGYYVQPTIFPGHNPRIFOEIEFGPVSVTSFDDFDVAKTAND 426
 QY 432 TOYGLGAGVWSRGNLAYKMGRIQACRVWNTNCYHAYFAHAFGGYKQSGIGRETHKMWL 491
 DB 427 TLYGLGAGVWTRMNTNAYRAGRALQACRVWNTNCYHAYFAHAFGGYKQSGIGRENKMWL 486
 QY 492 EHYQOTKCLLVSYSDKPLGLF 512
 DB 487 EHYQOTKCLLVSYSPKRLGLF 507

RESULT 8

ID DHAL_VIBCH STANDARD; PKT; 506 AA.
 AC P23240: Q9KTS0;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Aldehyde dehydrogenase (EC 1.2.1.3).
 GN ALDA OR VC0819.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCHI_TaxID-666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KPR 56.
 RX MEDLINE-91210174; PubMed-1902210;
 RA Parsot C., Mekalanos J.J.;
 RT "Expression of the Vibrio cholerae gene encoding aldehyde
 RT dehydrogenase is under control of ToxR, the cholera toxin
 RT transcriptional activator.";
 RL J. Bacteriol. 173:2842-2851(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-El Tor N16961 / Serotype O1;
 RX MEDLINE-98169509; PubMed-9501228;
 RA Karaolis D.K.R., Johnson J.A., Bailey C.C., Boedeker E.C., Kaper J.B.,
 RA Reeves P.R.;
 RT "A Vibrio cholerae pathogenicity island associated with epidemic and
 RT pandemic strains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:3134-3139(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Classical Ogawa 395 / ATCC 39541 / Serotype O1;
 RX MEDLINE-21117025; PubMed-11179381;
 RA Karaolis D.K.R., Tan R., Kaper J.B., Reeves P.R.;
 RT "Comparison of Vibrio cholerae pathogenicity islands in sixth and
 RT seventh pandemic strains.";
 RL Infect. Immun. 69:1947-1952(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-El Tor N16961 / Serotype O1;
 RX MEDLINE-20406833; PubMed-10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).

CC -1- FUNCTION: MAY BE INVOLVED IN V.CHOLERAE VIRULENCE, AS ITS
 CC EXPRESSION IS UNDER THE CONTROL OF TOXR, A TRANSCRIPTIONAL.

CC ACTIVATOR OF SEVERAL GENES ASSOCIATED WITH VIRULENCE.
 CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC -1- PATHWAY: Ethanol utilization; second step.
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

EMBL: M60658; AAA03051.1;
 EMBL: AF034434; AAC1274.1;
 EMBL: AF325733; AAK20747.1;
 EMBL: AE004167; AAF93982.1; ALT_INIT.
 DR HSSP: P05091; ICW3.
 DR TIGR: VC0819;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
 KW Oxidoreductase, NAD, Complete proteome.
 FT NP_BI: 240 245 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 262 262
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT ACT_SITE 401 401 BY SIMILARITY.
 SQ SEQUENCE 506 AA: 55884 MW: 698756 DGGFEL07 CRG64;

Query Match 62.9%; Score 1682; DB 1; Length 506;
 Best local similarity 63.4%; Prod. No. 1.1e-118;
 Matches 309; Conservative 67; Mismatches 112; Indels 0; Gaps 0;

QY 25 YDNFEGSEWVAPADAGYYUNLTHVTQULQLEVASSKKRDLALALAAUKUKIAHISVQ 84
 DB 19 YDNVIGGQWKKPHSGEYFSNISHVNGLVFCRVARSSQVVELADAAHNALESWSTISAV 78
 QY 85 DRAAILFKIADRMEONLELTAETWNGKPIRETSAADVPLAIDHRYFASCIQAQPG 144
 DB 79 ERSNILLRIADRIESNLETLAIVESDNGKPIRETIAALPLIDHRYFAAIAKSQGA 138
 QY 145 ISEVDSETVAYHPEHPEGVVQGIIDPNEPILMASKKMAIALAAGNIVLKVAKTIPESVL 204
 DB 149 ASLEDSKTLIYHUPPEHPEGVVQGIIDPNEPILMAAKRIALALAAAGIVLKIAEQIPVSTL 198
 QY 205 LLMEIVGDLPLPGVWVWVNVNAGCIVIGLEYLATSKRIAKVAFGTGTEWQQIDQYATONIIP 264
 DB 199 FLMEIIGDLIPAGVINNVNPGSEAGNALATSGRIDKLAFTGSTEIGNHILKCAADNLIP 258
 QY 265 VTLEIGGKSPNIVFAIVMDEHAFHFKALEGALFAPFNGCVCTCTSHALVQESIYERFM 324
 DB 259 SILEGGKSPNINIVFEHPEHPEGVVQGIIDPNEPILMAAKRIALALAAAGIVLKIAEQIPVSTL 418
 QY 325 FPAIPPVESIPSCNPIDSVTCMAQVSHQULETILNYIUIGKKEAD 484
 DB 319 AKILIERVALKQCNPLDTHIQCAQVSKQYQIKILYCIQIKDQDAELIFGHPHNNQNY 478
 QY 385 LKDGYYLEPTILFGQNNMRVFOEIEFGPVLAVTTEKTEEALAND 444
 DB 379 LSGGYIKPTLFGHNMHIFOEIEFGPVIAITKFKDELEALHLANITVYGLGAGVWTRD 438
 QY 445 GNLAYKMGKICAGRVWNTNCYHAYFAHAFGGYKQSGIGRETHKMWLEHYQTKILVSY 504
 DB 439 INTAHPMAKNIKAGPVWVWNTNCYHAYFAHAFGGYKQSGIGRETHKMWLEHYQTKILVSY 498
 QY 505 SIKPGLGF 512
 DB 499 EIHPLGLGF 506

RESULT 9

DHAM_MOUSE

ID DHAM_MOUSE STANDARD; PKT; 519 AA.

RL Arch Biochem Biophys. 289:214-223(1991).
 CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -|- PATHWAY: Ethanol utilization; second step.
 CC -|- SUBUNIT: HOMOTETRAMER.
 CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X14977; CAA33101.1; -;
 DR EMBL; M19030; AAA40719.1; -;
 DR PIR; A27113; A27113;
 DR PIR; S03564; S03564;
 DR PIR; S17492; S17492;
 DR HSSP; P05091; LCW3;
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldehyd; 1;
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_GYS; 1;
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1;
 KW Oxidoreductase; NAD; Acetylation; Mitochondrion; Transit peptide.
 FT TRANSIT 1 19 MITOCHONDRION.
 FT CHAIN 20 519 ALDEHYDE DEHYDROGENASE.
 FT MOD_RES 20 20 ACETYLATION (PROBABLE).
 FT NP_BIND 264 269 NAD (ADP PART) (BY SIMILARITY)
 FT ACT_SITE 287 287 BY SIMILARITY.
 FT ACT_SITE 321 321 BY SIMILARITY.
 SQ SEQUENCE 519 AA; 56488 MW; 75C748202F1333E5 CRC64;

 Query Match 36.5%; Score 975; DB 1; Length 519;
 Best Local Similarity 42.4%; Pred. No. 1.2e-65;
 Matches 208; Conservative 86; Mismatches 172; Indels 24; Gaps 10.

 QY 28 FIGEWAPADGVEYVNIPTVTQCLCEVASSGKRDIIDLAL---DAAHKVKDKWAHTSVQ 84
 DB 42 FINNEHDAVSKKTFPTVNSTGEVICVAEGNKEDVDKVAQAQAFAQLGSPWRMDAS 101

 QY 85 DRAALFKADREQNLLELATATWENCKPRTSAAADVPLAIDHFRFASCIRAQEGS 144
 DB 102 DRGLRLVRLADLERDRTYLAALLETLDNGKPYVISYLVLDLWLVLCRLRYAGWADKYHCK 161
 QY 145 ISEVDSITVAYHFEHPLGVVQGIIPWNFPLLMASKMAPALAAAGNCVILKPARLTPLSLV 204
 DB 162 TPIPDGFFSYTRHEPVGVAQGIIPWNFPLLMAWANKLGFALAGNVVWVKVAEQTLPLAL 221

 QY 205 LLMFTVQL-LPPGVVNVNAGAGVIGEYLATSKKIAKVAFITGTEVSGGIMUOYA-TUNI 262
 DB 222 VVANLKEAGFPFGVNVIVGFTGTAGATASHLDVDKVAFTGSTEVGHLLQVAAAGSNL 281

 QY 263 IPVTLEIGKSPNIVADVWDEDAFFDKALE--GFALFPNQGEVCTCPSPALVQESIIY 320
 DB 282 KRVTLLEGKSPNIMSE-----DAIMFWAVFQAIPALF-PNQGQCCAGSFTFVQEDVY 334

 QY 321 ERMERAIRVESIRSGNPLDSVTOMGAQVSHQLETTUNYIDIGKKEGADVLTGGRKL 380
 DB 335 DEEVSERVAKSRVVCNFTSPTETSPQVRETQPKILGYIKSQGEAKLLQGG----- 390

 QY 381 LEGELKD-GYLLKPTLLFG--QNNMKVFOFEIFGPIVAVTTPKTMHEALELANDTQYGLG 437
 DB 391 --GAADRGVFIQPTV-FGIVDKGMTIAKEPTEFGVPMQILKPKTFTEEVVGRANNSKYGA 447

 QY 438 AGVMSRGNLAYKMGKGTCAQPGWNTNCTHAYPAHAFAFGYKSGKIGCRETHKMMLFHYOOT 497
 DB 448 AAVTKULDKANVLSQALAGTVMINCYIVFCAQSPFGYKMGSGSRELPFGVFGIQAITEV 507

 QY 498 KCLLVSYSDK 507
 DB 508 KTVTVKVPQK 517

RESULT 11
 DHAM_HUMAN
 ID DHAM_HUMAN STANDARD; PRT; 517 AA.
 AC P05091; Q03639;
 DT 13-AUG-1987 (Rel. 05, Created)
 DI 01-JAN-1990 (Rel. 13, Last sequence update)
 FT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (E.C. 1.2.1.-.4) (ALDH
 DE class 2) (ALDH2) (ALDH1) (ALDH-E2).
 GN ALDH2 (OR ALDH).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCHI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE 88256152; PubMed 2848413;
 RA Hsu L.C., Iwendel R.E., Yoshida A.;
 RT "Genomic structure of the human mitochondrial aldehyde dehydrogenase
 RT gene.";
 RL Genomics 2:57-65(1988).
 RN [2]
 RP SEQUENCE OF 18-517.
 RC TISSUE=Liver;
 RX MEDLINE 86055846; PubMed 4065146;
 RA Hempel J., Kaiser R., Joernvall H.;
 RT "Mitochondrial aldehyde dehydrogenase from human liver. Primary
 RT structure, differences in relation to the cytosolic enzyme, and
 RT functional correlations.";
 RL Eur. J. Biochem. 153:13-28(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE 87174836; PubMed 3562250;
 RA Braun T., Huber E., Singh S., Agarwal D.P., Goedde H.W.;
 RT "Isolation and sequence analysis of a full length cDNA clone coding
 RT for human mitochondrial aldehyde dehydrogenase.";
 RL Nucleic Acids Res. 15:3179-3179(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE 87219091; PubMed 3582651;
 RA Braun T., Huber E., Singh S., Agarwal D.P., Goedde H.W.;
 RT "Evidence for a signal peptide at the amino terminal end of human
 RT mitochondrial aldehyde dehydrogenase.";
 RL FEBS Lett. 215:233-236(1987).
 RN [5]
 RP REVISIONS TO N-TERMINUS.
 RA Braun T., Huber E., Singh S., Agarwal D.P., Goedde H.W.;
 RL FEBS Lett. 233:440-440(1988).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg K.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 119-517 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE 85216574; PubMed 2987944;
 RA Hsu L.C., Tani K., Fujiyoshi T., Kurachi K., Yoshida A.;
 RT "Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4771-4775(1985).
 RN [8]
 RP SEQUENCE OF 119-517 FROM N.A.
 RX MEDLINE 85252089; PubMed 4015824;
 RA Yoshida A., Ikawa M., Hsu L.C., Tani K.;
 RT "Molecular abnormality and cDNA cloning of human aldehyde
 RT dehydrogenases.";
 RL Alcohol 2:104-106(1985).
 RN [9]
 RP SEQUENCE OF 214-500 FROM N.A.

TISSUE=Liver;
 MEDLINE=87279033; PubMed=3610592;
 RA Agarwal D.P., Goedde H.W.;
 RT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.";
 RL *Isotymes* Curr Top Biol Med Res 19:21-48(1987)
 RN [10]
 PP DES-REPTION OF ORIGIN OF GENELOCATIONS BETWEEN REF. 2 AND DNA SEQUENCES.
 RX MEDLINE=88005159; PubMed=3653404;
 RA Hempel J., Hoeog J.-O., Joernvall H.;
 RT "Mitochondrial aldehyde dehydrogenase: Homology of putative targeting
 sequence to that of carbamyl phosphate synthetase 1 revealed by
 correlation of cDNA and protein data.";
 PL *FEBS Lett* 222:95-98(1987)
 RN [11]
 RP VARIANT LYS-504.
 RX MEDLINE=84119449; PubMed=6582480;
 RA Yoshida A., Huang L.-Y., Ikawa M.;
 RT "Molecular abnormality of an inactive aldehyde dehydrogenase variant
 commonly found in Orientals";
 RL *Proc. Natl. Acad. Sci. U.S.A.* 81:258-261(1984)
 RN [12]
 RP VARIANT LYS-496.
 RX MEDLINE=96119362; PubMed=8561277;
 RA Noworadwsky A., Tsai S.J., Goldfarb E., Petersen P., Long J.C.,
 Goldman D.;
 RT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and
 American Indian populations: detection of new ALDH2 alleles.";
 RL *Alcohol Clin Exp Res* 19:1105-1110(1995)
 RN [13]
 RP X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).
 RX MEDLINE=20095857; PubMed=1061996;
 RA Ni L., Zhou Y., Hurley T.D., Weiner H.;
 RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional
 structure and the restoration of solubility and activity of chimeric
 forms.";
 PL *Protein Sci.* 8:2784-2790(1999).
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SUBUNIT: HOMOTETRAMER
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN
 ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS,
 CAN BE DIRECTLY ATTRIBUTED TO AN ENZYMATICALLY IMPAIRED ALDH
 ISOZYME. THE INACTIVE VARIANT (ALLELE 2 OR ALDH2*2) IS DUE TO
 SINGLE AMINO ACID EXCHANGE.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -!- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF. 1
 AND REF. 2 AND THAT IN REF. 3 AND 4 ARE PROBABLY ALL DUE TO
 FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF. 5 AND REF. 6.
 CC THE SEQUENCE DESCRIBED IN REF. 9 DIFFERS FROM THAT SHOWN DUE TO TWO
 FRAMESHIFTS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch)
 CC
 DR EMBL M20456; AAAS1693.1;
 DR EMBL M20444; AAAS1693.1; JOINED.
 DR EMBL M20445; AAAS1693.1; JOINED.
 DR EMBL M20446; AAAS1693.1; JOINED.
 DR EMBL M20447; AAAS1693.1; JOINED.
 DR EMBL M20448; AAAS1693.1; JOINED.
 DR EMBL M20449; AAAS1693.1; JOINED.
 DR EMBL M20450; AAAS1693.1; JOINED.
 DR EMBL M20451; AAAS1693.1; JOINED.
 DR EMBL M20452; AAAS1693.1; JOINED.
 DR EMBL M20453; AAAS1693.1; JOINED.
 DR EMBL M20454; AAAS1693.1; JOINED.
 DR EMBL X05409; CAA28940.1;

DR EMBL Y08109; CAA68290.1;
 DR EMBL W600267; AA02467.1;
 DR EMBL K03001; AA856500.1;
 DR EMBL M26769; AA81094.1;
 DR EMBL M43551; AA81825.1; ALT_FRAME.
 DR PIR: S00804; DEH0E2.
 DR PIR: A29975; A29975.
 DR PDB: 1CW3; 10-JAN-00.
 DR Genew: HGNC:404; ALDH2.
 DR MIM: 100650;
 DR InterPro: IPR02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR ProSITE: PS00670; ALDEHYDE_DEHYDR_CYS; 1.
 DR ProSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; Polymorphism;
 KW 3D-structure.
 FT TRANSIT 1 17 MITOCHONDRION
 FT CHAIN 18 517 ALDEHYDE DEHYDROGENASE.
 FT NF_BIND 262 267 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 285 285
 FT ACT_SITE 319 319
 FT VARIANT 337 337
 FT VARIANT 496 496
 FT VARIANT 504 504
 FT VARIANT 504 504
 FT CONFLICT 7 12 E -> V (IN DBSNP:1062136).
 FT CONFLICT 7 26 /FTID=VAR_011869.
 FT CONFLICT 18 18 E -> R (IN ALDH2*3).
 FT CONFLICT 40 85 E -> K (IN ALDH2*2); LPASTIC REDUCTION
 OF ENZYME ACTIVITY).
 FT CONFLICT 218 218 /FTID=VAR_002248.
 FT CONFLICT 247 247 PEGPPL -> APAPP (IN REF. 5).
 FT CONFLICT 340 340 PEGPPL -> APAPP (IN REF. 5).
 FT CONFLICT 462 463 PEGPPL -> APAPP (IN REF. 5).
 FT CONFLICT 517 517 PEGPPL -> APAPP (IN REF. 5).
 SQ SEQUENCE 517 AA; 56381 MW; E8F74D44D285A00E CRC64;

Query Match 46.38; Score 971; DB 1; Length 517;
 Best Local Similarity 42.74; Prod No 2 4e-65;
 Matches 209; Conservative 84; Mismatches 173; Indels 24; Gaps 10;
 QY 28 FIGGEVAPADGEYQNLTPVTGOLCEVASSKRDIDLALDAH---KVKDKWAHTSVQ 84
 PB 40 FINNEHVLAVSPKTPPVNNESTGEVIGVAPGKFLVDKAKAAPAFQIGSPWPMAS 99
 QY 85 LFAALIFKIALHMCNTELIATATPTWNGKPIETSAIVPLAIDHPPFASCTPAQFGS 144
 DB 100 HPGRIILNPLADLTPEDPTTAAITLNGKPYVTSIVTDIMVILKCTPYACWADKYHGK 159
 QY 145 ISEYDSETVAYHFEHPLGVGQIIPWNPFLMASWKMALAPALAGNCVWLKPARLTPLSVL 204
 DB 160 TPTPLQGFSTYTHPEWVGVGQIIPWNPFLMASWKMALAPALAGNCVWLKPARLTPLSVL 219
 QY 205 LIMEIVSEL-LFPGVW 252
 DB 220 YVANLTFAGPFGGWNIVPGPGTACAAIASHPIVWVWVWVWVWVWVWVWVWVWVWVWVW 279
 QY 283 LFTVLELGRKSNFIVFALVMLKELAFVDFALFEL---GHALFAFN---GHALVAFN 330
 DB 280 KKVTELEGRKSNFIVMSS-----HAILMVAWVWVWVWVWVWVWVWVWVWVWVWVWVW 332
 QY 321 ERFMEPALPKVESIFSNPLISVLMGACVSHSGLEPTILNYLIDGRKKEALVILNGPKKL 380
 DB 335 LEFVERSVAKAKSVW 388
 QY 381 LESELEKLD-GYYLPTILFG---QNNNKPVPQPEIFGPVLAVTIFKIMEEALANDJYGLG 437
 DB 389 --GIAADNGYPTQPTV-PSGVQDGMTAKKETEPVWMTLKPKTIFEVWVWVWVWVWVW 445
 QY 438 AGVSRNCLNAYKMGSPGIGQACPVWNTNYHAYPAHMAFGYGYKQSGKQSGPETHKMLRHYQCT 497


```

RP SEQUENCE OF 25-65.
RX MEDLINE=91113163; PubMed=1989592;
RA Chalmers R.M., Keen J.N., Fawson C.A.;
RT "Comparison of benzyl alcohol dehydrogenases and benzaldehyde
FT dehydrogenases from the benzyl alcohol and mandelate pathways in
RT Acinetobacter calcoaceticus and from the Tol-plasmid-encoded toluene
RT pathway in Pseudomonas putida. N-terminal amino acid sequences, amino
RT acid compositions and immunological cross-reactions.";
RL Biochem. J. 273:99-107(1991).
RN [3]
RP SEQUENCE OF 25-34 AND 378-386.
RX MEDLINE=97295278; PubMed=9150920.
RA Larsson T., Norbeck J., Karlsson H., Karlsson K.-A., Blomberg A.;
RT "Identification of two dimensional gel electrophoresis resolved yeast
RT proteins by matrix-assisted laser desorption/ionization mass
RT spectrometry.";
RL Electrophoresis 18:418-423(1997).
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=98340498; PubMed=9675847;
RA Tessier W.D., Meaden P.G., Dickinson F.M., Midgley M.;
RT "Identification and disruption of the gene encoding the K(+)-activated
RT acetaldehyde dehydrogenase of Saccharomyces cerevisiae.";
RL FEMS Microbiol. Lett. 164:29-34(1998).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z75282; CAA99705.1; -.
DR HSSP; P05091; ICW3.
DR SGD; S0005901; ALD4.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Mitochondrion; Transit peptide.
FT TRANSIT 1 24 MITOCHONDRION.
FT CHAIN 25 519 POTASSIUM-ACTIVATED ALDEHYDE
FT NP_BIND 268 273 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 290 290 POTENTIAL.
FT ACT_SITE 324 324 POTENTIAL.
FT CONFLICT 51 51 N -> NN (IN REF. 2).
FT CONFLICT 63 63 E -> V (IN REF. 2).
SQ SEQUENCE 519 AA; 56723 MW; E7D9944EA25F948E CRC64;

Query Match 35.2%; Score 940.5; DB 1, Length 519.
Best local similarity 41.4%; Pred. No. 4.7e-63;
Matches 205; Conservative 88; Mismatches 177; Indels 25; Gaps 10;

QY 18 PLKLGK--ARYDN-----FIGGEWVAPADGEYQYQNLTPVTGQLCEVASSCKRDIIDLALDAA 71
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 PIKLPNGLEVEQPTGLFINNKVPSPKQNTFEVINFSTEEICHIEGREDDVEFAVQAA 91
QY 72 HKV--KDKWHTSVQDRAALFKIADRMQONLELLATAETWNGKPIRETSAADVPLAID 129
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92 DRAFSNGSNWGIPIIDRGKALYPLAELIEQDKDVIATSIETLNGKAI--SSRGDWDIVIN 150
QY 130 HFRVFASCTRAQGGIISEVDSSETVAYHFEPLGVVCGQIIPNPFLLMASWMAFALAAAGN 189
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 YLKSSACFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPNPFLLMAWAKIAPALVTGN 210
QY 190 CVVLKPARLTPPLVILMEIVRGL-LPPGVVNVNAGSVIGEYLATSKRIAKVAFGTST 248
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 TVVLKTAESTPLSALYVSVIPDQGIIPGVINIVSGFGKIVGGEAITNHPKIKKVAFGTST 270

```

```

QY 249 EVGQOIMQYATONLIPVTLLELACKSPNIVFAIVMDEEDAFFUKALEGVAL FAFWQGEWC 407
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
271 AIGRHITQSAAALKKVILFLGCKSPNIVA--DAELKKAQVNIILGLIYNSGEWC 424
QY 308 TPTSPALVDSEIYEPFMERAIFFVESIPSNIPFISVLMCAVSRQLFTILNYILGCKK 467
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 CAVSKYVVESTYDKPIFFPFAASESKVDTTFDLSLPGAGLSUMGLNNILKYVVLGKN 484
QY 368 EGALVLTGGRRKILLEGELKINGVYLPFTLPG QNNMRVLDGDEFGPVLAVITTKMEFA 425
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 EEAIIITGEEI GSAVFETDVF FGVKFLMEIVKFFIEGVVIVKFAIADEV 438
QY 426 LELANDFOYGLGAGVWSKNNILAYKKMKGLQAGKQWNTNYHAYPAHAAGKQKSGTIGRE 485
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 INMANUSEYGLAAGIHTSINTALKVADRVNACTVWINTYNDFHIAVTFGGFNASGLGRE 498
QY 486 TRKMILPHYQOTKCL 500
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
499 MSVALQNYLQVKAV 514

```

Search completed: June 24, 2003, 10:14:31
Job time : 12.3122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:08:15 ; Search time 39.1863 Seconds
(without alignments)
2692.168 Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAGIKRQVGFPLK... HYGLKLLVSYSDKPLQLFSL

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacterioph.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2651	99.2	542	16 Q8XDJ1	Q8XDJ1 escherichia
2	2601	97.3	512	16 Q8ZL72	Q8ZL72 salmonella
3	2589	96.9	512	16 Q8ZL72	Q8ZL72 salmonella
4	1948	72.9	502	16 Q8ZL70	Q8ZL70 rhizobium m
5	1918	71.8	506	16 Q8ZL71	Q8ZL71 rhizobium m
6	1905	71.3	505	16 Q8ZL71	Q8ZL71 rhizobium m
7	1895	70.9	540	16 Q8ZL71	Q8ZL71 rhizobium m
8	1879	70.4	505	2 Q8ZL71	Q8ZL71 rhizobium m
9	1878	70.3	506	16 Q8ZL71	Q8ZL71 rhizobium m
10	1870	70.0	506	2 Q8ZL71	Q8ZL71 rhizobium m
11	1850.5	69.2	503	2 Q8ZL71	Q8ZL71 rhizobium m
12	1850.5	69.2	506	2 Q8ZL71	Q8ZL71 rhizobium m
13	1848.5	64.2	505	16 Q8ZL71	Q8ZL71 rhizobium m
14	1802	67.4	506	16 Q8ZL71	Q8ZL71 rhizobium m
15	1791	67.0	506	16 Q8ZL71	Q8ZL71 rhizobium m
16	1791	67.0	506	16 Q8ZL71	Q8ZL71 rhizobium m

17 1789 66.9 506 2 Q8ZAA1
18 1781 66.6 506 2 Q8ZAA1
19 1780.5 66.6 506 2 Q8ZAA1
20 1774 66.4 506 2 Q8ZAA1
21 1772 66.3 506 16 Q8XU08
22 1756 65.7 506 2 Q8ZAA1
23 1755 65.7 506 2 Q8ZAA1
24 1754 65.6 506 2 Q8ZAA1
25 1257.5 47.0 491 16 Q8R663
26 1246.5 46.6 495 16 Q8ZAA1
27 1052 39.4 499 2 Q8XU08
28 1038 38.8 494 2 Q8R663
29 1032 38.6 496 16 Q8ZAA1
30 1021.5 38.2 537 5 Q8ZAA1
31 1010 37.8 501 2 Q8ZAA1
32 1008.5 37.7 511 2 Q8ZAA1
33 986 36.9 495 16 Q8ZAA1
34 972 36.4 501 2 Q8ZAA1
35 970 36.3 503 3 Q8ZAA1
36 969.5 36.3 507 10 Q8ZAA1
37 963.5 36.0 510 5 Q8ZAA1
38 958.5 35.9 494 3 Q8ZAA1
39 954 35.7 516 13 Q8ZAA1
40 954 35.7 511 13 Q8ZAA1
41 953.5 35.7 511 10 Q8ZAA1
42 953 35.7 518 13 Q8ZAA1
43 952 35.6 518 13 Q8ZAA1
44 952 35.6 519 11 Q8ZAA1
45 949 35.5 520 5 Q8ZAA1

ALIGNMENTS

RESULT 1

Q8XDJ1 PRELIMINARY: PRI: 542 AA.

ID Q8XDJ1

AC Q8XDJ1

DT 01-MAR-2002 (TREMREL. 20, Created)

DT 01-MAR-2002 (TREMREL. 20, Last sequence update)

DT 01-JUN-2002 (TREMREL. 21, Last annotation update)

DE Alderhyde dehydrogenase B (lactaldehyde dehydrogenase).

GN ALDH B 75008 (P05446).

OS Escherichia coli O157:H7.

OC Bacteria, Proteobacteria, gamma subdivision, Enterobacteriaceae.

OC Escherichia.

OX NCBI_TaxID=83334;

RN [1]

PP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose P.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamianou K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch P.A., Blattner F.P.

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7; Nature 403:529-533(2001).

EN [2]

EE SEQUENCE FROM N.A.

PC STRAIN=O157:H7 / FIMD 0509952;

EX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Ohnishi K., Ohnishi M., Murata T., Tanaka M., Taka T., Han C.-G., Chikuba E., Nakayama K., Nakayama C., Ogasawara N., Yasunaga T., Tada T., Takami H., Honda T., Sasakawa C., Shinagawa H., Kohara S., Shiba T., Hattori M., Shinagawa H.

RA "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12; PNA Res 8:11-22(2001).

DR EMBL: AF005585; AAG58731.1;

DR EMBL: AF005566; BAB37887.1;

DP InterPro: IPR020066; Alderhyde_dehydr

```

DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Complete proteome.
SQ SEQUENCE 542 AA; 59/68 MW; FA08/AA4BEEY02A70 CRC64;

Query Match 99.2%; Score 2651; DB 16; Length 542;
Best Local Similarity 99.2%; Pred. No. 1.4e-198;
Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGYGFPLKARYUNFICGGEWAPADGEYYONIIPVTGQLICEVASSG 60
DL 1 MTNPPSAQIKPGYGFPLKARYUNFICGGEWAPADGEYYONIIPVTGQLICEVASSG 90
QY 61 KRDIIDALDAHKVKOKWAHTSVQDRAAIIKFIADMEQNIPLATAETWNGKPIRETS 120
DB 91 KRDIIDALDAHKVKOKWAHTSVQDRAAIIKFIADMEQNIPLATAETWNGKPIRETS 150
QY 121 AADVPLAIDHFRVFASCIKRAQEGGISEVDSSETVAYHFHEPLGVGQIIPWNEPPLMASWK 180
DB 151 AADVPLAIDHFRVFASCIKRAQEGGISEVDSSETVAYHFHEPLGVGQIIPWNEPPLMASWK 210
QY 181 MAPALAGNCVILKPARLTPLSVLLMEIVGDLIPGVVNVVNGAGGIVGEYLATSKRIA 240
DB 211 MAPALAGNCVILKPARLTPLSVLLMEIVGDLIPGVVNVVNGAGGIVGEYLATSKRIA 270
QY 241 KVAFTGSTVGQOIMQYATONIIIPVTELGKSPNIFADVMDEDAFDKALEGFALFA 300
DB 271 KVAFTGSTVGQOIMQYATONIIIPVTELGKSPNIFADVMDEDAFDKALEGFALFA 330
QY 301 FNGEVCICPSKALVQESIVERMERAIRFVSIESRGNPLISVIMIAVSNHSGLETTIN 360
DB 331 FNGEVCICPSKALVQESIVERMERAIRFVSIESRGNPLISVIMIAVSNHSGLETTIN 390
QY 361 YIDIGKKGADVLITGGRKLEGEELKGYLLEPTILFGNNNRVPOEEIFGVLAIVTFK 420
DB 391 YIDIGKKGADVLITGGRKLEGEELKGYLLEPTILFGNNNRVPOEEIFGVLAIVTFK 450
QY 421 TMEALELANETOYGLGAGVWSRNGNLAYKMGKGIQAGKVTNICYHAYPAHAAFGYKQS 480
DB 451 TMEALELANETOYGLGAGVWSRNGNLAYKMGKGIQAGKVTNICYHAYPAHAAFGYKQS 510
QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
DB 511 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 542

RESULT 2
ID Q8ZL72 PRELIMINARY; PRT: 512 AA.
AC Q8ZL72;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Aldehyde dehydrogenase B (lactaldehyde dehydrogenase) (EC 1.2.1.22).
DE 1.2.1.22).
GN ALDH OR SIM3680.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OX NCBI_TaxID=602;
FN [1]
PP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SCS1412 / ATCC 700720.
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Tayman B.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston K., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).

```

```

DR EMBL: AE008870; AAL22539.1;
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 512 AA; 56/126 MW; 770189E05E5G87DA CRC64;

Query Match 97.8%; Score 2601; DB 16; Length 512;
Best Local Similarity 96.8%; Pred. No. 1.4e-194;
Matches 494; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGYGFPLKARYUNFICGGEWAPADGEYYONIIPVTGQLICEVASSG 60
DB 1 MTNPPSAQIKPGYGFPLKARYUNFICGGEWAPADGEYYONIIPVTGQLICEVASSG 60
QY 61 KRDIIDALDAHKVKOKWAHTSVQDRAAIIKFIADMEQNIPLATAETWNGKPIRETS 120
DB 61 KRDIIDALDAHKVKOKWAHTSVQDRAAIIKFIADMEQNIPLATAETWNGKPIRETS 120
QY 121 AADVPLAIDHFRVFASCIKRAQEGGISEVDSSETVAYHFHEPLGVGQIIPWNEPPLMASWK 180
DB 121 AADVPLAIDHFRVFASCIKRAQEGGISEVDSSETVAYHFHEPLGVGQIIPWNEPPLMASWK 180
QY 181 MAPALAGNCVILKPARLTPLSVLLMEIVGDLIPGVVNVVNGAGGIVGEYLATSKRIA 240
DB 181 MAPALAGNCVILKPARLTPLSVLLMEIVGDLIPGVVNVVNGAGGIVGEYLATSKRIA 240
QY 241 KVAFTGSTVGQOIMQYATONIIIPVTELGKSPNIFADVMDEDAFDKALEGFALFA 300
DB 241 KVAFTGSTVGQOIMQYATONIIIPVTELGKSPNIFADVMDEDAFDKALEGFALFA 300
QY 301 FNGEVCICPSKALVQESIVERMERAIRFVSIESRGNPLISVIMIAVSNHSGLETTIN 360
DB 301 FNGEVCICPSKALVQESIVERMERAIRFVSIESRGNPLISVIMIAVSNHSGLETTIN 360
QY 361 YIDIGKKGADVLITGGRKLEGEELKGYLLEPTILFGNNNRVPOEEIFGVLAIVTFK 420
DB 361 YIDIGKKGADVLITGGRKLEGEELKGYLLEPTILFGNNNRVPOEEIFGVLAIVTFK 420
QY 421 TMEALELANETOYGLGAGVWSRNGNLAYKMGKGIQAGKVTNICYHAYPAHAAFGYKQS 480
DB 421 TMEALELANETOYGLGAGVWSRNGNLAYKMGKGIQAGKVTNICYHAYPAHAAFGYKQS 480
QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
DB 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512

RESULT 3
ID Q8Z2D5 PRELIMINARY; PRT: 512 AA.
AC Q8Z2D5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Aldehyde dehydrogenase B (EC 1.2.1.22).
GN STY4116.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OX NCBI_TaxID=601;
FN [1]
PP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.P., Thomson N.R., Pickard D., White J.,
RA Chicher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies K.M., Dowd L., White N., Farrar J.,
RA Holtwell L., Hamlin N., Haque A., Hien T.T., Holroyd S., Jancels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Ratberford K., Simmonds M., Skellern T., Struelens K.,

```

RA Whitehead S., Barrell B.G.,
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT *enterica* serovar Typhimurium CT18.",
 RI Nature 413:848-852(2001).
 RL EMBL: AL627281; CAL07945.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR OXIDOREDUCTASE; Complete proteome.
 KW SEQUENCE 512 AA; 56355 MW; A61CD3C0FA7E394D CRC64;
 SQ
 Query Match 96.9%; Score 2589; DB 16; Length 512;
 Best Local Similarity 95.7%; Pred. No 9,2e-194;
 Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MTNPPSAQIKPGEYGEPLKARYDNFIDGEWVAPADGEYQNLTPVTGQLLCEVASSG 60
 DB 1 MTNPPSPRTIQSEYGYPLKARYDNFIDGEWVAPADGEYQNLTPVTGQLLCEVASSG 60
 QY 61 KRDI DALDAAHKVKWAHTSVQDRAAILFKIADRMENLELLATAETWNGKPIRETS 120
 DB 61 KKLULALGAHAKAKIWAHTSVQDRAAILFKIADRMENLELLATAETWNGKPIRETS 120
 QY 121 AADVPLADHRYFASCIQAEVGGISEVDSSETVAYHFEPLGVVGQIIPWNPFLLMASWK 180
 DB 121 AADVPLADHRYFASCIQAEVGGISEVDSSETVAYHFEPLGVVGQIIPWNPFLLMASWK 180
 QY 181 MAPALAGNCVVLKPARLTPLSVLLLMFIVGDLPLPGVNVVNGAGGIVGEYLATSKRTA 240
 DB 181 MAPALAGNCVVLKPARLTPLSVLLLMFIVGDLPLPGVNVVNGAGGIVGEYLATSKRTA 240
 QY 241 KVAFTGSTVEVGGQIMQYATONIIPVTLFGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 DB 241 KVAFTGSTVEVGGQIMQYATONIIPVTLFGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 QY 301 FNGEVCCTCPSRALVQESIIYEPFNERAIPRVESIPSONPLDSVTQMGAVSHQLETLIN 360
 DB 301 FNGEVCCTCPSRALVQESIIYEPFNERAIPRVESIPSONPLDSVTQMGAVSHQLETLIN 360
 QY 361 YIDICKKGKGAIVLTGGRRKILFGELEKPYLYLEPTIFPGNNMRVFEIIPGVLAVTTTK 420
 DB 361 YIDICKKGKGAIVLTGGRRKILFGELEKPYLYLEPTIFPGNNMRVFEIIPGVLAVTTTK 420
 QY 421 TMEALELANDIQVGLGAGVWSRNCNIAYKMGRIQAGRVWTCYHAYPAHAAPGGYKQS 480
 DB 421 TMEALELANDIQVGLGAGVWSRNCNIAYKMGRIQAGRVWTCYHAYPAHAAPGGYKQS 480
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 DB 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512

RESULT 4
 Q92N70 PRELIMINARY; PRT; 502 AA.
 ID Q92N70
 AC Q92N70; (TREMELrel, 19, Created)
 DT 01-DEC-2001 (TREMELrel, 19, Last sequence update)
 DT 01-JUN-2001 (TREMELrel, 21, Last sequence update)
 DE Probable aldehyde dehydrogenase protein (E: 2.1.1.3)
 GN R02351 (P. SWC02669)
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capella D., Parley-Hubler F., Gouzy T., Rothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu F., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss F., Lelaure V., Masuy D.,

RA Pohl T., Pottelie D., Puchler A., Pottelie B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Calbert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT *Sinorhizobium meliloti* strain 1021.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL: AL591790; CAC46930.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR OXIDOREDUCTASE; Complete proteome.
 KW SEQUENCE 502 AA; 55250 MW; D45A262260REB436 CRC64;
 SQ
 Query Match 72.9%; Score 1948; DB 16; Length 502;
 Best Local Similarity 73.5%; Pred. No 1.1e-143;
 Matches 361; Conservative 51; Mismatches 79; Indels 0; Gaps 0;
 QY 22 KARYDNFIDGEWVAPADGEYQNLTPVTGQLLCEVASSGKPIDLALDAAHKVKWAHT 81
 DB 12 KARYDNFIDGEWVAPADGEYQNLTPVTGQLLCEVASSGKPIDLALDAAHKVKWAHT 81
 QY 82 SVQDRAAILFKIADRMENLELLATAETWNGKPIRETSADVPLADHRYFASCIQAE 141
 DB 72 SVQDRAAILFKIADRMENLELLATAETWNGKPIRETSADVPLADHRYFASCIQAE 141
 QY 142 EGGISEVDSSETVAYHFEPLGVVGQIIPWNPFLLMASWKMAPALAGNCVVLKPARLTP 201
 DB 132 EGGISEVDSSETVAYHFEPLGVVGQIIPWNPFLLMASWKMAPALAGNCVVLKPARLTP 201
 QY 202 SVLLMEIVGDLPLPGVNVVNGAGGIVGEYLATSKRTAETWNGKPIRETSADVPLADHRYFASCIQAE 261
 DB 192 SVLLMEIVGDLPLPGVNVVNGAGGIVGEYLATSKRTAETWNGKPIRETSADVPLADHRYFASCIQAE 261
 QY 262 IIPVTLFGKSPNIFADVMDDEDAFFDKALEGFALFAFNGEVCCTCPSRALVQESIIY 321
 DB 252 IIPVTLFGKSPNIFADVMDDEDAFFDKALEGFALFAFNGEVCCTCPSRALVQESIIY 321
 QY 322 REMERAIPRVESIPSONPLDSVTQMGAVSHQLETLINIDICKKGKGAIVLTGGRRKIL 381
 DB 312 REMERAIPRVESIPSONPLDSVTQMGAVSHQLETLINIDICKKGKGAIVLTGGRRKIL 381
 QY 382 FGELEKPYLYLEPTIFPGNNMRVFEIIPGVLAVTTTKTMEALELANDIQVGLGAGV 441
 DB 372 FGELEKPYLYLEPTIFPGNNMRVFEIIPGVLAVTTTKTMEALELANDIQVGLGAGV 441
 QY 442 SNGENLAYKMGRIQAGRVWTCYHAYPAHAAPGGYKQSGIGRETHKMMLEHYQOQTK 501
 DB 432 SNGENLAYKMGRIQAGRVWTCYHAYPAHAAPGGYKQSGIGRETHKMMLEHYQOQTK 501
 QY 502 VSYSDKPLGLF 512
 DB 492 VSYSDKPLGLF 502

RESULT 5
 Q9A211 PRELIMINARY; PRT; 506 AA.
 ID Q9A211
 AC Q9A211; (TREMELrel, 17, Created)
 DT 01-JUN-2001 (TREMELrel, 17, Last sequence update)
 DT 01-MAR-2002 (TREMELrel, 20, Last annotation update)
 DE Aldehyde dehydrogenase B.
 GN CC3580.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.F.,
 RA Eisen J., Heidelberg J.F., Alley M.P.K., Ohta N., Muddock J.P.,

RA Potocka L., Nelson W.C., Newton A., Stephens C., Phadke N.D., Fly R.,
 RA DeBoy R.T., Dodson P.L., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolony J.F., Smir J., Craven M.R., Khouri H., Sherry J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Frimolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of *Caulobacter crescentus*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AF006016; AAK25542.1; ..
 DR HSP: P05091; ICW3.
 DR TIGR: CC3580; ..
 DR InterPro: IPRO02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 506 AA: 55661 MW: 268909ALE526050 CRG64.

Query Match 71.8%; Score 1918; DB 16; Length 506;
 Best Local Similarity 71.7%; Pred. No. 2.4e-141;
 Matches 352; Conservative 57; Mismatches 82; Indels 0; Gaps 0;
 QY 22 KARYDNFICGEWAPAGEYONLTPTVTGOLCEVASSCKRDLALDAAHKVKKWAHT 81
 DB 16 KARYDNFICGEWAPAGEYONLTPTVTGOLCEVASSCKRDLALDAAHKVKKWAHT 75
 QY 82 SVQDRAAILEKIAIDRMFONLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIRAQ 141
 DB 76 SAADRSILLRIADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIRAQ 135
 QY 142 EGGISEVDSTVAYHHEPIGVVGQIIPWNPFLMASWKMALAGNCVVLKPARLTP 201
 DB 136 EGGISEVDSTVAYHHEPIGVVGQIIPWNPFLMASWKMALAGNCVVLKPARLTP 195
 QY 202 SVLLIMEIVGDLPPGVVNVVNGAGVIGEXYLAETSKRIAKVAFTGSIWVGQIIMYATON 261
 DB 196 SIMWAEIMIGDLPPGVVNVVNGAGVIGEXYLAETSKRIAKVAFTGSIWVGQIIMYATON 255
 QY 262 IIPVTLELGKSPNIFADWDEDAFFDKALEGAFNAGVCTCPSPRALVOESIVE 321
 DB 256 IIPVTLELGKSPNIFADWDEDAFFDKALEGAFNAGVCTCPSPRALVOESIVE 315
 QY 322 KEMERALKRNVAVVQSPDPATMCAQAFSEQLNKLIGYMDIGNFGAKLARGQRKIL 381
 DB 316 KEMERALKRNVAVVQSPDPATMCAQAFSEQLNKLIGYMDIGNFGAKLARGQRKIL 375
 QY 382 EGELKDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTTEKMEALELANDTOYGLGAGW 441
 DB 376 PGQLADGVYVEPTVFECHNKRIFQEEIFGPVLAVTTTEKMEALELANDTOYGLGAGW 435
 QY 442 SRNCLAYKMGRIQACRVNTNCHAYPAHAFAFGYKQSGIGRETHKMMLEHYQOTKCLL 501
 DB 436 SRNCLAYKMGRIQACRVNTNCHAYPAHAFAFGYKQSGIGRETHKMMLEHYQOTKCLL 495
 QY 502 VSYSDKPIGLF 512
 DB 496 VSYSPRALGPF 506

RESULT 6
 Q98806
 ID Q98806 PRELIMINARY; PRT; 505 AA.
 AC Q98806
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Aldehyde dehydrogenase
 GN ML6639.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Mesorhizobium.
 OX NCBI_TaxID:381;
 RN [1]
 RP SEQUENCE FROM N.A.

STRAIN-MAFF403099;
 MEDLINE:21082940; PubMed:11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu F., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa G., Kohata M., Matsumoto M., Matsumoto A.,
 RA Morishizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti,"
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP004009; BAB52891.1; ..
 DR InterPro: IPRO02086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 505 AA: 55586 MW: 158763AFAPFA37 CRG64;
 Query Match 71.4%; Score 1905; DB 16; Length 505;
 Best Local Similarity 70.9%; Pred. No. 2.5e-140;
 Matches 352; Conservative 63; Mismatches 84; Indels 2; Gaps 1;
 QY 14 EYGFPLK--LKARYDNFICGEWAPAGEYONLTPTVTGOLCEVASSCKRDLALDA 71
 DB 5 EFSRPVKAPDKRYKNGFICGKWTPEKSGRYFENHSPVNGQLCEVARSALADIEAALDA 64
 QY 72 HKYDKWAHTSVQDRAAILEKIAIDRMFONLELLATAETWNGKPIRETSAADVPLAIDH 141
 DB 65 HAAKDAWGTSTVAFPSLILNPIADRMFONLELLATAETWNGKPIRETSAADVPLAIDH 124
 QY 132 RYFASCIKRAOEGGISEVDSTVAYHHEPIGVVGQIIPWNPFLMASWKMALAGNCV 191
 DB 125 RYFASAVPQEGSLSQIHQTVAYHHEPIGVVGQIIPWNPFLMASWKMALAGNCV 184
 QY 192 VTPAPLILPLSVLLIMEIVGDLPPGVVNVVNGAGVIGEXYLAETSKRIAKVAFTG 251
 DB 185 VLKPAQTAAIILWADLIGDLPPGVVNVVNGAGVIGEXYLAETSKRIAKVAFTG 244
 QY 252 QCLMYATONIIIPVTLELGKSPNIFADWDEDAFFDKALEGAFNAGVCTCPSPRAL 411
 DB 245 KLMYASQNLIPVTLELGKSPNIFADWDEDAFFDKALEGAFNAGVCTCPSPRAL 404
 QY 312 PAIVQESTVERPMEPAIPVVESTIPSCNPLDSVTOMCAQVSHQIPTIINYIDKPKK 471
 DB 305 PAIVQESTVERPMEPAIPVVESTIPSCNPLDSVTOMCAQVSHQIPTIINYIDKPKK 464
 QY 372 VLTGGRKRLLEGKDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTTEKMEALELAND 431
 DB 365 VLTGGRNVLPGDLGAGYVYKPTVFGHKNMIFQEEIFGPVSVVTTTEKMEALELAND 424
 QY 432 TOYGLGAGVWSRNGNLIAYKMGRIQACRVNTNCHAYPAHAFAFGYKQSGIGRETHKMM 491
 DB 425 TLYGLGAGVWSRNGNLIAYKMGRIQACRVNTNCHAYPAHAFAFGYKQSGIGRETHKMM 484
 QY 492 EHYQOTKCLLVSYSQKPIGLF 512
 DB 485 DRYQOTKCLLVSYSQKPIGLF 505

RESULT 7
 Q98806
 ID Q98806 PRELIMINARY; PRT; 505 AA.
 AC Q98806
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Aldehyde dehydrogenase
 GN ALDA GR AT02224 OR AGR_C_4041.
 OS Aerobacterium tumefaciens (strain C58 / ATCC 34970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID:176299;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21508550; PubMed=11743193;
 PA Wood P.W., Serubai T.C., Kaul P., Monks P.F., Kitajima T.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D.S.,
 RA Chapman P., Clendenning J., Weatherall G., Gillet W., Grant C.,
 PA Kutyavin T., Levy R., Li M.-J., McLelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenghimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tan Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tinley S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 PT C58.";
 PL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Roumel K., Gordon J., Vaubin M., Tatchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 PT Agrobacterium tumefaciens C58.";
 PL Science 294:242-253(2001).
 DR EMBL: AF009172; AAL43213.1; ALT_INIT.
 DR EMBL: AF008138; AAK7966.1;
 KW Complete proteome.
 SQ SEQUENCE 540 AA; 59646 MW; B03074B38B1A8E46 CRC64;

Query Match 70.9%; Score 1895; DB 16; Length 540;
 Best Local Similarity 69.0%; Pred. No. 1.7e-139;
 Matches 348; Conservative 68; Mismatches 84; Indels 4; Gaps 1;

QY 9 QIKPGYFPLKAKYDNFPGGFWAPADGGYYQNI:TPVTGQLLCEVASSGKPIDLAL 68
 DB 41 QOKAGEAPFKL---YGYNGKWEKPGSGPYMDNI:SPVTGKHKIFVPPSDASDTEPAL 96
 QY 69 DAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSAADVPLAI 128
 DB 97 DAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSAADVPLAI 156
 QY 129 DFRYFASCIRAQEGGISEVDSETVAYHFHEPLGAVGQIIPWNPFLMASWKMAPALAA 188
 DB 157 DFRYFASCIRAQEGGISEVDSETVAYHFHEPLGAVGQIIPWNPFLMASWKMAPALAA 216
 QY 189 NPVVIKPAELIPLSVILLMEIVGILLPGVNVVNGAGVGLGYLATSKEIAVFTGST 248
 DB 217 NPVVIKPAELIPLSVILLMEIVGILLPGVNVVNGAGVGLGYLATSKEIAVFTGST 276
 QY 249 EVGQITMGVATQNTIPVTLELGGKSPNLFVAVMUEEALFUKALEGFALFAPNAGEVCT 308
 DB 277 SVGKELMYAADNTNISLELGGKSPNLFVAVMUEEALFUKALEGFALFAPNAGEVCT 336
 QY 309 CPSRALVQESYVERPMEKATIRVFTSRGNPLDSVTOMGACVSHGQLETLINIDIGKKE 368
 DB 337 CPSRALVHESYIDREMEKAIKKVQAISQDDPLNPSTNLGASQAEQFDKMSYLEIGKKE 396
 QY 369 GAVVIGGFKKLEELKGYIIPVTLEFGNNMPVETKIPSPVLAVTFKIMVETFKIM 428
 DB 397 GAVVIGGFKKLEELKGYIIPVTLEFGNNMPVETKIPSPVLAVTFKIMVETFKIM 456
 QY 429 ANDVYGLGAGVWMSKNGLAYKMGKGIGAGVWNTNCHAYPAHAAGFGYKQSGIGRETHK 488
 DB 457 ANDVYGLGAGVWMSKNGLAYKMGKGIGAGVWNTNCHAYPAHAAGFGYKQSGIGRETHK 516
 QY 489 MMEHYQOTKCLLVSYSDKPIGLF 512
 DB 517 MMEHYQOTKCLLVSYSDKPIGLF 540

RESULT 8
 Q56203
 ID 056203 PPEIIMINAPY; PPT: 505 AA
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Chloroacetaldehyde dehydrogenase-
 GN ALDA.
 OS Xanthobacter autotrophicus
 OC Plasmid linear plasmid pXAO1.
 OC Bacteria, Proteobacteria; alpha subdivision, Rhizobiaceae group;
 OC Hyphomicrobium group; Xanthobacter.
 OX NCBI_TaxID=280;
 RN [1]
 PP SEQUENCE FROM N.A.
 RX MEDLINE=98172729; PubMed=9511738;
 RA Bergeron H., Iahbe D., Turmel C., Lau P.C.;
 RT "Cloning, sequence and expression of a linear plasmid based and a
 RT chromosomal homolog of chloroacetaldehyde dehydrogenase-encoding genes
 RT in Xanthobacter autotrophicus GJ10.";
 RL Gene 207:9-18(1998).
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DP FMRL; AF029733; AAC13641.1;
 DR HSP; P05091; ICM3.
 DE 100-100%; IP0002846; A3dhy3d_dehyd
 DR Prtm: PF00171; aldehyd; 1.
 DE PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1
 DE PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; plasmid.
 SQ SEQUENCE 505 AA; 54945 MW; AD6A9EDB0297BB04 CRC64;

Query Match 70.3%; Score 1879; DB 2; Length 505;
 Best Local Similarity 69.2%; Pred. No. 2.7e-138;
 Matches 353; Conservative 58; Mismatches 93; Indels 6; Gaps 1;

QY 3 NNPIPSAQIKPGYFPLKAKYDNFPGGFWAPADGGYYQNI:TPVTGQLLCEVASSGKPI 62
 DB 2 NKPEIATIKQSPF-----KAPYGNFTGGKVEPVGPGPYFNT:SPVTGKHKIFVPPSDAD 55
 QY 63 DIDALDAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSA 122
 DB 56 DIEKALDAHAAKAWGKTSAAVRALILNKLADPMEONLELLATAETWNGKPIRPIETSA 115
 QY 123 DVPALDAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSA 182
 DB 116 DVPALDAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSA 175
 QY 183 DVPALDAHKKVKQKWAHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSA 242
 DB 176 PALAAGNCWLKPAEQTPASILVLAELIADLLPPGVNINVGLEAGKPLASSSTIAKI 245
 QY 243 APTSTSTVGGQIMQVATQNTIPVTLELGGKSPNLFVAVMUEEALFUKALEGFALFAPN 402
 DB 236 APTSTSTVGGQIMQVATQNTIPVTLELGGKSPNLFVAVMUEEALFUKALEGFALFAPN 295
 QY 303 QSEVETDPSKALVQESYVERPMEKATIRVFTSRGNPLDSVTOMGACVSHGQLETLINID 362
 DB 296 QSEVETDPSKALVQESYVERPMEKATIRVFTSRGNPLDSVTOMGACVSHGQLETLINID 355
 QY 363 GAVVIGGFKKLEELKGYIIPVTLEFGNNMPVETKIPSPVLAVTFKIMVETFKIM 422
 DB 356 DVGKQGAEEVLIGGEPNTFEGDLAGYGVYVVKPTFKGHKKRIFQGFTEFGVWVTFETD 415
 QY 423 FEALRELANTQYGLGAGVWMSKNGLAYKMGKGIGAGVWNTNCHAYPAHAAGFGYKQSGI 482
 DB 416 FEALRELANTQYGLGAGVWMSKNGLAYKMGKGIGAGVWNTNCHAYPAHAAGFGYKQSGI 475
 QY 483 GPETHKMLHRYQOTKCLLVSYSDKPIGLF 512
 DB 476 GPETHKMLHRYQOTKCLLVSYSDKPIGLF 505


```

10 Q9PDS1 PRELIMINARY; PRT; 503 AA.
AC Q9PDS1.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aldehyde dehydrogenase 1.
GN ALD1.
OS Acinetobacter sp. M-1
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=123502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M-1.
RX MEDLINE=20378657; PubMed=10919810;
RA Ishige T., Tani A., Sakai Y., Kato N.;
RT "Long-chain aldehyde dehydrogenase that participates in n-alkane
RT utilization and wax ester synthesis in Acinetobacter sp. strain M-1.";
PI Appl Environ Microbiol. 66:3481-3486(2000)
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL; AB042203; BAB11888.1;
DR HSSP; P05091; 1CW3.
DR InterPro; IPR003086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 503 AA; 55499 MW; 3254DBACCC65F246 CRC64;

Query Match 69.2%, Score 1850.5, DB 2, Length 503,
Best Local Similarity 67.6%, Pred. No. 4.5e-136;
Matches 340; Conservative 70; Mismatches 90; Indels 3; Gaps 1;

QY 10 IKPEYSEPLKAKRYDNFTGGWVAFADGEYYQNLIPTVGLLCEVASSGKKUIGLALD 69
DB 4 VDPNCGSK IHFKQYENF IGGWVAPVKYVDNISPVDGKSFTIRPSSARDIELAD 63

QY 70 AAHKVKQKWAHTSVQDRAALFKIADRMEOLELLATAETDNGKPIRTSAADPLAID 129
DB 64 AAHKAKKEWKKSSPTTRSNLLKTIADRMLEANLEMAVETDNGKPVRETLAID 123

QY 130 HFRYFASCIQAEGGISEVDSEYVAYHFHEPLGVGQIIPWNPPLLMASWKPALAAGN 189
DB 124 HFRYFACIQAEGGISEIDEETIAYHFHEPLGVGQIIPWNPPLLMASWKPALAAGN 183

QY 130 CWLKPARELTPSLVLLMEIVGDLPPGVVNVNAGVGTGEYLATSKRIKVAFTGSTE 249
DB 184 CWLKPARELTPVGLLVALLVAILQDILPAGVLINVNGYGAEGRPLATSPRIKIATGTQ 243

QY 250 VGGQIMQYATNTITFTVLELGGKSPNIVFALVWMELEAFEDKALESPALFAFNQKRVTC 409
DB 234 VGLIMQYATNTITFTVLELGGKSPNIVFALVWMELEAFEDKALESPALFAFNQKRVTC 403

QY 310 PSRALVQESIVEREMERAIIRVESIRSNPLDVSVMCAVSHGJLETLILNYIDGKKKPS 469
DB 304 PSRALVQESIVEREMERAIIRVESIRSNPLDVSVMCAVSHGJLETLILNYIDGKKKPS 463

QY 370 ADVLTGRRKLLLEGELKDGYYLEPTILFGQNNMRVFOEEIFGVLAVTTFF 429
DB 364 AEVLGGHGR---QEVNGYVLEPTIFKGNHNMVFOEEIFGVLAVTTFFKDFDPAQTQ 420

QY 430 NDTQYCLGAGVWSNGNLAYKMGRTGAGRVWNTCYHAYPAHAAGGYSQSGIRFTHKM 489
DB 421 NDTMYGLGAGVWSNSTHATYAGRAIIPAGSPVWNTCYHAYPAHAAGGYSQSGIRFTHKM 480

QY 480 MIEHYQOTKGLVSYSPKPLGLF 512
DB 481 MLDHYQOTKLLVSYSTKMGFF 503

RESULT 12
O50206
1D O50206 PRELIMINARY; PRT; 506 AA.

```

```

O50206;
01-JUN-1998 (TrEMBLrel. 06, Created)
01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE NAD-dependent chloroacetaldehyde dehydrogenase.
GN ALDB.
OS Xanthobacter autotrophicus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Xanthobacter.
OX NCBI_TaxID=280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GJ10.
RX MEDLINE=98172729; PubMed=9511738;
RA Bergeron H., Labbe D., Turmel C., Lau P.C.;
RT "Cloning, sequence and expression of a linear plasmid-based and a
RT chromosomal homolog of chloroacetaldehyde dehydrogenase-encoding genes
RT in Xanthobacter autotrophicus GJ10.";
PI Gene 207:9-18(1998).
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY
DR EMBL; AF029734; AAC13644.1;
DR HSSP; P05091; 1CW3.
DR InterPro; IPR003086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 506 AA; 54885 MW; C92B41F49C38AAB7 CRC64;

Query Match 69.2%, Score 1850.5, DB 2, Length 506;
Best Local Similarity 66.5%, Pred. No. 4.6e-136;
Matches 341; Conservative 71; Mismatches 90; Indels 11; Gaps 2;

QY 3 NNP---PSAQLKPEYGFPLKLRKYDNFTGGWVAFADGEYYQNLIPTVGLLCEVASS 59
DB 2 NSPAFLISATSP-----FKRYGNFTGGTFFVPSGPPYFNTSPITRKVFIEVPS 53

QY 60 GKPTIDTALDAAHKVKQKWAHTSVQDRAALFKIADRMEOLELLATAETDNGKPIRT 119
DB 54 DEHDTERALDAHAARKKAWGPTSAADPAIPLVLIKIADPLEANLELLAGAEETWNGKPIRT 113

QY 120 SAALVPLIALHFRYFASCIQAEGGISEVDSEYVAYHFHEPLGVGQIIPWNPPLLMASW 179
DB 114 TLADTPLAIDHFRYFACVRAQEGTILGEIDHDIIAYHYHPLGVGQIIPWNPPLIMAAW 173

QY 180 KMAFALAAGNVTVLKIAFLTPHSLVLLMEIVGDLPPGVVNVNAGVGTGEYLATSKRI 249
DB 174 KIAPAGAGNVVLKPAFOIPASILLVLAELIADLLPPGVLINVNGYGAEGRPLATSPRI 233

QY 240 AKVAFIGSIEVGGQIMQYATNTITFTVLELGGKSPNIVFALVWMELEAFEDKALESPALF 299
DB 234 SKIAFTGEIGTRELINGYAAQNLIPTVLELGGKSPNIPFADVAEDDEFLKAVEGVWF 293

QY 300 AFNLTQVCTGSPALVQESIVEREMERAIIRVESIRSNPLDVSVMCAVSHGJLETLIL 359
DB 294 ALNCEVCTGSPALVQESIVEREMERAIIRVESIRSNPLDVSVMCAVSHGJLETLIL 353

QY 360 NYIDIGKKEGADVLTGRRKLLLEGELKDGYYLEPTILFGQNNMRVFOEEIFGVLAVTTFF 419
DB 354 SYIDIGRAEQVLAGGFRNVLTGELADGFYVKTPTFLCKNNMRIFQEEIFGVLAVTTFF 413

QY 420 KTMFEALFELANDYVGGAGVWSNPNINLAYKMKRPITLACGKVVWNTCYHAYPAHAAGGYSQ 479
DB 414 KQTFEALFIANCTLFGLGAGVWSNPNINLAYKMKRPITLACGKVVWNTCYHAYPAHAAGGYSQ 473

QY 480 SGIGRETHKMMLEHYQOTKGLVSYSPKPLGLF 512
DB 474 SGIGRETHKMMLEHYQOTKGLVSYSPKPLGLF 506

RESULT 13
O50206
1D O50206 PRELIMINARY; PRT; 505 AA.

```